
Konzepte und Guidelines für Applikationen in
Cinematic Virtual Reality

DISSERTATION

an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

eingereicht von
Diplom-Mathematiker

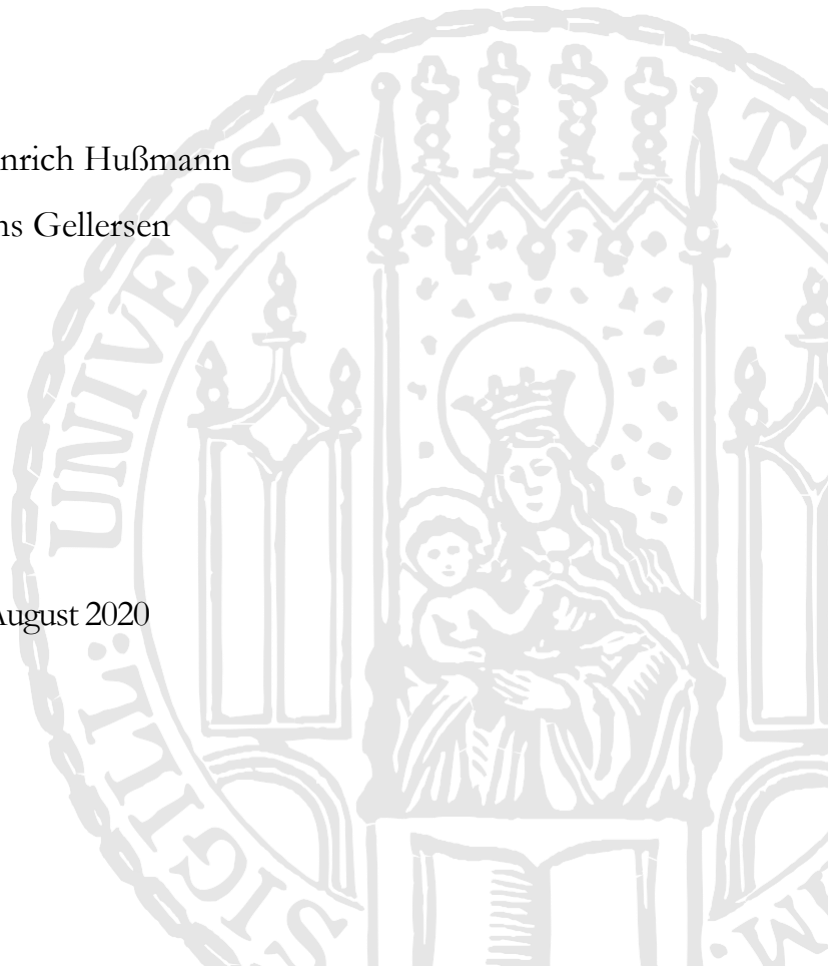
Sylvia Rothe

München, den 23. April 2020



1. Gutachter: Prof. Dr. Heinrich Hußmann
2. Gutachter: Prof. Dr. Hans Gellersen

Tag der mündlichen Prüfung: 5. August 2020



Zusammenfassung

Die meisten Menschen, die zum ersten Mal einen omnidirektionalen Film über ein Head-Mounted Display (HMD) sehen, sind fasziniert von der neuen Erlebniswelt. Das Gefühl, an einem anderen Ort zu sein, weit weg von der Realität, beeindruckt und lässt sie in eine andere Welt eintauchen. Die über Jahrzehnte entwickelte Filmsprache lässt sich nicht ohne Weiteres auf dieses neue Medium - Cinematic Virtual Reality (CVR) - übertragen. Der Betrachter kann die Blickrichtung und damit den sichtbaren Ausschnitt des Bildes frei wählen, und es ist deshalb nicht immer möglich, dem Zuschauer zu zeigen, was für die Geschichte wichtig ist. Herkömmliche Methoden für die Lenkung der Aufmerksamkeit - wie Nahaufnahmen oder Zooms - sind nicht ohne Weiteres verwendbar, andere - wie Bewegung und Farben - benötigen eine Evaluation und Anpassung. Um neue Konzepte und Methoden für CVR zu finden, sind neben den Forschungsergebnissen aus dem Filmbereich auch die anderer Forschungsgebiete, wie Virtual und Augmented Reality (VR und AR), relevant.

Um geeignete Techniken der Aufmerksamkeitslenkung in CVR zu ergründen, werden in dieser Arbeit bekannte Methoden aus Film, VR und AR analysiert und eine einheitliche Taxonomie präsentiert. Dadurch ist es möglich, die verschiedenen Aspekte detaillierter zu untersuchen.

Auch die Positionierung der Kamera kann nicht ohne Weiteres vom traditionellen Film auf CVR übertragen werden. Der Zuschauer nimmt bei der Betrachtung einer CVR-Anwendung in der virtuellen Welt die Position der Kamera ein. Dies kann zu Problemen führen, wenn die Kamerahöhe nicht seiner eigenen Körpergröße entspricht. Außerdem ist eine Auflösung einer Szene durch verschiedene Einstellungsgrößen nicht ohne Weiteres möglich, da dies für den Zuschauer ein Umherspringen in der virtuellen Welt bedeuten würde. In dieser Arbeit werden die Auswirkungen verschiedener Kamerapositionen auf den Zuschauer untersucht und Guidelines zur Kamerapositionierung vorgestellt.

Die dazugewonnene Raumkomponente bietet neue Möglichkeiten. Schnitte müssen nicht unbedingt von der verstrichenen Zeit abhängen, sondern können auch auf der Blickrichtung des Betrachters basieren. In Übereinstimmung mit dem Begriff Timeline führen wir das Konzept der Spaceline für diese Methode der Story-Konstruktion ein. Während die Schnitte auf der Timeline vom Filmmacher festgelegt werden, bestimmt der Betrachter die Spaceline - innerhalb eines vom Filmmacher festgelegten Konstrukts. Durch diese individuelle Zuschauerführung ist es möglich, dass jeder seine eigene Geschichte in seinem eigenen Tempo und mit seinen eigenen Prioritäten entdeckt.

Das Spaceline-Konzept bietet neue Interaktionsmöglichkeiten, die durch verschiedene Selektionstechniken umgesetzt werden können. Um Techniken zu finden, die für CVR geeignet sind, werden in dieser Arbeit blick- und kopfbasierte Ansätze untersucht. Auch wenn deren Wirksamkeit stark von den gewählten Parametern und physiologischen Faktoren abhängen, konnten wertvolle Erkenntnisse gewonnen werden, die in einen Design-Space für CVR-Konstrukte einfließen. Dieser Design-Space ermöglicht es beim Entwerfen einer CVR-Anwendung, die Attribute zu finden, die für die Anwendung am besten geeignet sind.

Aber nicht nur die Entwicklung von CVR-Anwendungen stellt neue Herausforderungen. Durch das HMD ist ein Zuschauer von der restlichen Welt isoliert, und es bedarf neuer Methoden, um CVR als soziale Erfahrung erlebbar zu machen. Einige davon werden in dieser Arbeit vorgestellt und analysiert. Aus den gewonnenen Erfahrungen werden Empfehlungen für einen CVR-Movie-Player abgeleitet.

Um die vorgestellten Konzepte und Guidelines zu entwickeln, wurden eine Reihe von Nutzerstudien durchgeführt, zum Teil mit Aufzeichnung der Kopf- und Blickrichtungen. Um diese Daten analysieren zu können, wurde ein Tool entwickelt, welches die Visualisierung der Daten auf dem Film ermöglicht.

In dieser Arbeit werden Konzepte und Guidelines für verschiedene Felder in Cinematic Virtual Reality vorgestellt: Aufmerksamkeitslenkung, Kamerapositionierung, Montage, Zuschauererlebnis und Datenanalyse. Auf jedem dieser Gebiete konnten Erkenntnisse gewonnen werden, die auch für die andere Bereiche von Interesse sind. Oft hängen die Ergebnisse der einzelnen Fachgebiete zusammen und ergänzen sich gegenseitig. Ziel der Arbeit ist es, die verschiedenen Aspekte als Ganzes zu präsentieren.

Abstract

Most people who watch an omnidirectional film for the first time on a head-mounted display (HMD) are fascinated by the new world of experience. The feeling of being in a different place, right in the middle of the action, far from reality, impresses and gives the opportunity to immerse in another world. The film language developed over decades cannot simply be transferred to this new media, Cinematic Virtual Reality (CVR). The viewer can freely choose the direction of view and thus the visible section of the picture and it is therefore not always possible to show the viewer what is important for the story. Traditional methods for directing attention - such as close-ups or zooms - are not easy to use, others - such as movement and colors - needs to be assessed and adjusted. For finding new concepts and methods for CVR, in addition to the research results from the film area, other areas in the field of virtual and augmented reality (VR and AR) are relevant.

In order to find suitable methods to draw attention in CVR, known methods from film, VR, and AR are analyzed in this work and a uniform taxonomy is presented. This makes it possible to investigate the various aspects of the methods in more detail.

The positioning of the camera cannot simply be transferred from traditional film to CVR. When viewing a CVR application in the virtual world, the viewer takes the place of the camera in the real world. This can lead to problems if the camera height does not match the viewer's height. In addition, a resolution of a scene due different setting sizes is not possible as this would mean that the viewer would jump around in the virtual world. In this work, the effects of different camera positions on the viewer are examined and guidelines for camera positioning are presented.

The additional space component offers new possibilities. Cuts do not necessarily have to depend on the elapsed time, but can also be based on the viewer's gaze. In accordance with the term timeline, we introduce the concept of the spaceline for this method of story plot construction. While the cuts on the timeline are determined by the filmmaker, the viewer determines the spaceline - within a construct determined by the filmmaker. Through this individual guided tour, everyone can discover their own story at their own pace and with their own priorities.

The spaceline concept offers new interaction options that can be implemented using various selection techniques. In order to find methods that are suitable for CVR, this work examines eye and head-based techniques. Even if the effectiveness of them strongly depends on the chosen parameters and physiological factors, valuable insights are gained, which are included in a design space for spaceline constructs. This design space allows one to find the attributes that best suits for the application when designing a CVR application.

But not only the creation of CVR applications presents new challenges. The HMD isolates a viewer from the rest of the world and requires new techniques to experience CVR in a social way. Some of these are presented and analyzed in this work. Recommendations for a CVR movie player are derived from the experience gained.

Several user studies were conducted to develop the concepts and guidelines, some of them by recording of the head and gaze directions. To be able to analyze this data, a tool was developed which enables us to visualize the data on the film.

In this work, concepts and guidelines for various fields in Cinematic Virtual Reality are presented: attention guiding, camera positioning, montage, audience experience, and data analysis. In each of these areas, knowledge was gained that is also of interest to the other fields. The findings of the individual fields are often related and complement each other. This work aims to present the various aspects as a whole.

Disclaimer

Diese Arbeit ist das Ergebnis mehrerer Projekte, die in den letzten drei Jahren durchgeführt wurden. Alle Projekte waren nur mit der Unterstützung von Kollegen und Studenten möglich. Das Folgende ist eine Übersicht darüber, wer an den Projekten beteiligt und was mein eigener Beitrag zu jedem dieser Projekte war.

Kapitel 2

Dieses Kapitel enthält bereits veröffentlichtes Material aus mehreren Publikationen mit den Co-Autoren Heinrich Hußmann, Daniel Buschek, Mohamed Khamis, Felix Althammer und Mathias Allary (Rothe u. a., 2019a, 2018a, 2017; Rothe und Hußmann, 2018). Die Inhalte wurden gemeinsam mit den Co-Autoren diskutiert, die Publikationen von mir verfasst und das Feedback der Co-Autoren eingearbeitet. Felix Althammer führte die Studien zu (Rothe u. a., 2018a) durch und wurde dabei von Mohamed Khamis und mir betreut. Die Studie zu (Rothe und Hußmann, 2018) wurde von mir selbst durchgeführt.

Die zu diesem Kapitel gehörenden Studien aus Abschnitt 7.1 basieren auf folgenden studentischen Abschlussarbeiten, welche von mir betreut wurden - die Bachelorarbeit von Felix Althammer gemeinsam mit Mohamed Khamis. Die Studien wurden von den Studenten durchgeführt und jeder Projektschritt in regelmäßigen Treffen gemeinsam besprochen. Die Daten wurden von mir neu analysiert und interpretiert.

- Duc Anh Le (2017): Nondiegetic Methods for Guiding Attention in Cinematic Virtual Reality (Bachelor)
- Ziheng Gao (2018): Gaze Direction by High Frequency Flickering in Cinematic Virtual Reality (Master)
- Felix Althammer (2018): Guiding the User's Gaze in 360 Degree Videos (Bachelor)

Kapitel 3

Dieses Kapitel enthält bereits veröffentlichtes Material aus mehreren Publikationen mit den Co-Autoren Heinrich Hußmann, Vanessa Sarakiotis, Boris Kegeles, und Mathias Allary (Rothe u. a., 2019b, 2019d, 2018d) Die Inhalte wurden gemeinsam mit den Co-Autoren diskutiert, die Publikationen von mir verfasst und das Feedback der Co-Autoren eingearbeitet. Vanessa Sarakiotis führte die Studien zu (Rothe u. a., 2019d) durch und wurde dabei von mir betreut. Die Studien zu (Rothe u. a., 2019b, 2018d) wurde unter meiner Betreuung von Boris Kegeles durchgeführt.

Die zu diesem Kapitel gehörenden Studien aus Abschnitt 7.2 basieren auf folgenden studentischen Abschlussarbeiten, welche von mir betreut wurden. Die Studien wurden von den Studenten durchgeführt und jeder Projektschritt in regelmäßigen Treffen gemeinsam besprochen. Die Daten wurden von mir neu analysiert und interpretiert.

- Pia Probst (2020): Proxemics in Cinematic (Bachelor)
- Vanessa Sarakiotis (2018): Space Composition in Cinematic Virtual Reality (Bachelor)
- Boris Kegeles (2017): Influences of Camera Position to Attention and Presence in CVR (Bachelor)
- Nicole Gora (2019): Cuts and Camera Movements in Cinematic Virtual Reality (Bachelor)

Kapitel 4

Dieses Kapitel enthält bereits veröffentlichtes Material aus mehreren Publikationen mit den Co-Autoren Heinrich Hußmann, Daniel Buschek und Harald Brunner (Rothe u. a., 2019c, 2018c, 2018b; Rothe und Hußmann, 2019). Die Inhalte wurden gemeinsam mit den Co-Autoren diskutiert, die Publikationen von mir verfasst und das Feedback der Co-Autoren eingearbeitet. Harald Brunner führte die Studien zu (Rothe u. a., 2018b, 2018c) unter meiner Betreuung durch.

Die zu diesem Kapitel gehörenden Studien aus Abschnitt 7.3 basieren auf folgenden studentischen Abschlussarbeiten, welche von mir betreut wurden. Die Studien wurden von den Studenten durchgeführt und jeder Projektschritt in regelmäßigen Treffen gemeinsam besprochen. Die Daten wurden von mir neu analysiert und interpretiert.

- Evelyn Müller, (2018): Differences Between Cinematic VR and Traditional Movies (Bachelor)
- Harald Brunner (2018): Affordances for Interaction in Cinematic Virtual Reality via WebVR (Master)
- Pascal Pothmann, (2019): Activation Methods in Cinematic Virtual Reality (Bachelor)
- Dominik Beyer (2019): Camera Positions in Cinematic Virtual Reality (Bachelor)

Kapitel 5

Dieses Kapitel enthält bereits veröffentlichtes Material aus mehreren Publikationen mit den Co-Autoren Heinrich Hußmann, Kim Tran, Mario Montagud, Christian Mai und Daniel Buschek (Rothe u. a., 2018f, 2018g, 2018e). Die Inhalte wurden gemeinsam mit den Co-Autoren diskutiert, die Publikationen von mir verfasst und das Feedback der Co-Autoren eingearbeitet. Kim Tran führte die Studien zu (Rothe u. a., 2018f, 2018g) unter meiner Betreuung durch.

Die zu diesem Kapitel gehörenden Studien aus Abschnitt 7.4 basieren auf folgenden studentischen Abschlussarbeiten, welche von mir betreut wurden.

- Alexander Schmidt (2018): Collaborator Awareness in Cinematic Virtual Reality (Bachelor)
- Viet Duc Ngo (2018): The Influence of Viewport Sharing on Social Awareness in CVR (Bachelor)
- Vo Duc Dung (2018): Methods of Communication for Social Viewing in Cinematic VR (Bachelor)
- Atef Azabi (2018): Influence of Voice Chat for Social Viewing in Cinematic VR (Bachelor)
- Lang Zhao (2018): Less than 360° in Cinematic Virtual Reality (Master)
- Arne Fahrenwalde (2019): Less than 360° in Cinematic Virtual Reality (Bachelor)
- Kim Tran (2017): Subtitles in Cinematic Virtual Reality (Bachelor)
- Lili Tran (2017): Embedding of Sign Language in Cinematic VR (Bachelor)
- Lisa Görtz (2017): Methods of Speaker Identification for Subtitles in Cinematic VR (Bachelor)
- Michaela Schmierl (2018): Enhanced Subtitle Methods in Cinematic Virtual Reality (Bachelor)

Die Studien wurden von den Studenten durchgeführt und jeder Projektschritt in regelmäßigen Treffen gemeinsam besprochen. Die Daten wurden von mir neu analysiert und interpretiert.

Danksagung

Die vorliegende Dissertation hätte ohne die Unterstützung zahlreicher Personen in dieser Form nicht realisiert werden können. An erster Stelle gilt mein Dank meinem Doktorvater Heinrich Hußmann für seine außerordentliche wissenschaftliche und methodische Unterstützung während der gesamten Dissertationsphase. Danke für die vielen konstruktiven Anregungen, die richtigen Fragen zum richtigen Zeitpunkt und die Möglichkeiten, mich mit Wissenschaftlern vieler Länder austauschen zu können. Diese internationale Zusammenarbeit war eine große Bereicherung für meine Arbeit, und gemeinsame Veröffentlichungen sind dabei entstanden. Besonderer Dank gilt dabei Tobias Höllerer, der ein Jahr an unserem Institut in München arbeitete und mir wertvolle Ratschläge für meine wissenschaftliche Arbeit gab; und auch meinen Kollegen in Barcelona, Valencia und Lancaster für den Erfahrungsaustausch über Cinematic Virtual Reality, insbesondere Mario Montagud für die vielen Diskussionen zu Social Viewing. Besondere Dank gilt dabei Hans Gellersen für seine wertvollen Anregungen und die hilfreichen Gespräche.

Danke auch an Mathias Allary für die zahlreichen Gespräche über kinematische Aspekte und dafür, dass er seinen großen filmischen Erfahrungsschatz mit mir teilte. Ein großes Dankeschön auch an Martin Rieger der mir mit seinem Wissen über spatial Sound jederzeit weiterhelfen konnte.

Für die vielfältige Unterstützung bin ich besonders dem Deutschen Museum München verbunden, in dem wir mehrere Nutzerstudien durchführen durften. Speziell bedanken möchte ich mich bei Andrea Geipel, Claus Henkensiefken und Alexander Schmidt für die hervorragende Zusammenarbeit bei den gemeinsam durchgeführten Projekten.

Vielen Dank auch an das Institut für Rundfunktechnik und der Arbeitsgemeinschaft Behinderung und Medien für die wertvollen Gespräche zu barrierefreien Medien und der Unterstützung der Nutzerstudien in diesem Gebiet. Insbesondere möchte ich mich bedanken bei Markus Ludwig, Hermann Höbel und Simon von der Au für die ausgezeichnete Zusammenarbeit in dieser Zeit.

Bedanken möchte ich mich auch bei den Bachelor- und Masterstudenten, die ich in dieser Zeit betreuen durfte und ohne die diese Arbeit so nicht möglich gewesen wäre. Die große Anzahl an Nutzerstudien, die wir in dieser Zeit durchführen konnten, um Erfahrungen auf dem Gebiet Cinematic Virtual Reality zu sammeln, haben wesentlich zu dieser Dissertation beigetragen. Deshalb vielen Dank an: Duc Anh Le, Kim Tran, Evelyn Müller, Vanessa Sarakiotis, Pia Probst, Alexander Schmidt, Pascal Pothmann, Michaela Schmierl, Harald Brunner, Simon von der Au, Felix Althammer, Boris Kegeles, Ziheng Gao, Lang Zhao, Luca Speeter, Arne Fahrenwalde, Nicola Gora, Dominik Beyer, Dung Vo Duc, Atef Azabi, Viet Duc Ngo, Lisa Görtz, Lilli Tran, Richard Schäffer, Michaela Finger und Niklas Esch.

Meinen wissenschaftlichen Kollegen am Institut für Medieninformatik der LMU danke ich für die immer zielführenden Diskussionen. Speziell danken möchte ich dabei: Daniel Buschek, Ceenu George, Mohamed Khamis, Beat Rossmys, Heiko Drewes und Lewis Chuang. Ebenso geht mein Dank an Rainer Fink, Franziska Schwamb und Christa Feulner, die stets dafür sorgten, dass alle Arbeitsabläufe gut funktionierten und für Probleme immer ein offenes Ohr hatten.

Ich möchte Eric Sprigg dafür danken, dass er in der Anfangszeit meiner Dissertation jede englischsprachige Veröffentlichung detailliert korrigierte und mir dadurch den Start erst ermöglichte.

Besonders dankbar bin ich meiner Familie, die meine Arbeit mit großem Engagement und viel Geduld unterstützt hat. Großer Dank gebührt meinem Mann Detlef Rothe für seine moralische Unterstützung, seine Unterstützung bei der Implementierung des CVR-Analyzers und das unermüdliche Korrekturlesen in dieser Zeit. Vielen Dank auch an meinen Enkel Matheo Rothe, der genau an dem Tag zur Welt kam, an dem ich mit dieser Dissertation begann. Wir lernten in dieser Zeit unsere neue Welt verstehen und an ihr teilzuhaben - jeder von uns beiden auf seine Weise. Mit seinem Lachen und seinen Fragen holte er mich immer wieder zurück in die reale Welt und zeigte mir wie schön sie ist.

Inhaltsverzeichnis

Vorwort.....	1
Herausforderungen.....	1
Wissenschaftliche Beiträge	2
Struktur der Arbeit.....	2
Stilistische Entscheidungen.....	4
Abkürzungen	5
1 Einleitung.....	6
1.1 Historischer Rückblick	7
1.1.1 Vom bewegten Bild zum Film	7
1.1.2 Panorama-Filme, 360°-Filme und Sphärische Filme	8
1.1.3 VR und AR	10
1.1.4 Cinematic VR.....	10
1.2 Was ist Cinematic VR?	11
1.2.1 Begriffe	11
1.2.2 Vergleich von framed Filmen und Cinematic VR	11
1.2.3 Vergleich zwischen Game und Cinematic VR	13
1.2.4 Vergleich zwischen Real-time VR und Cinematic VR	14
1.3 Anwendungsgebiete, Möglichkeiten und Grenzen von CVR.....	14
1.4 CVR als Forschungsgebiet.....	16
2 How to Guide: Aufmerksamkeitslenkung in CVR	17
2.1 Hintergrund	19
2.1.1 Aufmerksamkeitstheorie / Attention Theory	19
2.1.2 Grundlagen der Physiologie der Augen	20
2.1.3 Aufmerksamkeitslenkung im Film	21
2.1.4 Guiding Methoden.....	22
2.2 Taxonomie.....	25
2.2.1 Diegetisch und nicht-diegetisch	26
2.2.2 Visuell, auditiv und haptisch.....	26
2.2.3 On- und Off-Screen.....	27
2.2.4 World- und Screen-referenziert	28
2.2.5 Direkte und indirekte Cues	29
2.2.6 Subtil und apparent	29
2.2.7 System-forciert, reflex-forciert und freiwillig	30
2.3 Verwendung der Taxonomie	30
2.3.1 Guiding Methoden in framed Filmen und Bildern für CVR.....	31
2.3.2 Guiding Methoden in VR und AR für CVR	33
2.3.3 Designüberlegungen bei der Anwendung der Taxonomie	34
2.4 Fazit und Ausblick.....	36

3	How to Shoot: Kameraführung in CVR	37
3.1	Theoretischer Hintergrund	38
3.1.1	Regeln der Kameraführung	38
3.1.2	Proxemik	42
3.1.3	Sichtfeld - Blickfeld - Gesichtsfeld	43
3.1.4	Augenhöhe	45
3.2	Kamerahöhe in CVR	46
3.2.1	Diskrepanz in der Körpergröße	46
3.2.2	Diskrepanz in Postures	47
3.2.3	Fazit	48
3.3	Kamerapositionen in CVR	49
3.3.1	Kameraabstände und Einstellungsgrößen in CVR	49
3.3.2	F-Formationen in CVR	51
3.3.3	Fazit	51
3.4	Kamerabewegungen in CVR	52
3.4.1	Kamerafahrten in CVR	52
3.4.2	Fazit	53
3.5	Guidelines für Kameraführung in CVR	53
3.6	Fazit und Ausblick	54
4	How to Cut: Montage in CVR mittels Spaceline-Konzept	55
4.1	Theoretischer Hintergrund	56
4.1.1	Psychologische Aspekte	58
4.1.2	Schnitt via Timeline	58
4.1.3	Interaktive Narration	61
4.1.4	Interaktion	63
4.2	Spaceline-Konzept	64
4.2.1	Terminologie	65
4.2.2	Selektionstechniken in CVR	67
4.2.3	Transitions: Montagetechniken in CVR	70
4.3	Realisierung von Spaceline-Konstrukten	73
4.3.1	Guidelines für die Konstruktion einer Spaceline-Applikation	73
4.3.2	Design-Space für Spaceline Konstrukte	75
4.3.3	Konzept für einen CVR-Editor	77
4.4	Fazit und Ausblick	82
5	How to Watch: Nutzererlebnis in CVR	83
5.1	Gemeinsame CVR-Anwendungen	85
5.1.1	Theoretischer Hintergrund	85
5.1.2	Herausforderungen für Social Viewing in CVR	87
5.1.3	Methoden für kollaborative CVR-Erlebnisse	87
5.1.4	Folgerungen für die Herausforderungen	93

5.1.5	Design-Space für Social Viewing Anwendungen.....	94
5.1.6	Weiterführende Konzepte und Ausblick.....	97
5.2	Komfortable CVR-Anwendungen	98
5.2.1	Theoretischer Hintergrund	99
5.2.2	Herausforderungen für angenehme CVR-Applikationen.....	101
5.2.3	Methoden zur Gestaltung komfortabler CVR-Applikationen	102
5.2.4	Folgerungen für die Herausforderungen	105
5.3	Barrierefreie CVR- Anwendungen	105
5.3.1	Theoretischer Hintergrund	106
5.3.2	Herausforderungen für barrierefreie CVR-Experiences.....	108
5.3.3	Methoden für barrierefreie CVR-Experiences	108
5.3.4	Folgerungen für die Herausforderungen	111
5.3.5	Weiterführende Konzepte und Ausblick.....	111
5.4	Realisierung von CVR-Player-Applikationen.....	112
5.4.1	Design-Space für CVR-Player-Anwendungen.....	112
5.4.2	Konzept für einen CVR-Player.....	113
5.5	Fazit und Ausblick.....	116
6	How to Evaluate: Analysemethoden für CVR.....	117
6.1	Theoretischer Hintergrund.....	118
6.2	CVR-Analyzer: Ein Tool zum Analysieren von Cinematic VR-Tracking-Daten.....	120
6.2.1	Anforderungen	121
6.2.2	Implementierung.....	121
6.2.3	Zusammenfassung.....	128
6.3	Räumliche Statistik	129
6.3.1	Raumzeitwürfel	129
6.3.2	Getis-Ord (Gi*) Statistik.....	130
6.4	Fragebögen für Cinematic VR.....	131
6.4.1	Presence	131
6.4.2	Task Workload	132
6.4.3	Simulator-Sickness.....	132
6.4.4	User Experience und Enjoyment	132
6.5	Fazit und Ausblick.....	132
7	Nutzerstudien.....	133
7.1	Nutzerstudien: Aufmerksamkeitslenkung.....	134
7.1.1	Studie: Diegetisches Guiding	134
7.1.2	Studie: Vergleich spatial Sound – nonspatial Sound.....	145
7.1.3	Studie: Guiding mittels PiP oder Fading	149
7.1.4	Studie: Guiding durch Flicker.....	152
7.1.5	Studie: Guiding durch subtile Flicker und Eyetracking	155
7.1.6	Studie: Guiding durch hochfrequente Flicker	159

7.2	Nutzerstudien: Kameraführung.....	161
7.2.1	Studie: Proxemiks und Kameraabstände	161
7.2.2	Studie: F-Formationen	165
7.2.3	Studie: Diskrepanzen in der Körpergröße.....	173
7.2.4	Studie: Diskrepanzen in der Körperhaltung.....	180
7.2.5	Studie: Kamerabewegungen.....	182
7.3	Nutzerstudien: Montage	185
7.3.1	Studie: Richtungswechsel bei Transitions.....	186
7.3.2	Studie: Loops vor Transitions.....	189
7.3.3	Studie: On-Screen Indikatoren: Vergleich von Cursor und Shape	192
7.3.4	Studie: Off-Screen Indikatoren: Vergleich von Linien und Kantenspot.....	195
7.3.5	Studie: Blickbasierte Interaktionstechniken	197
7.3.6	Studie: Kameraausrichtungen beim Spacewechsel	201
7.4	Nutzerstudien: Nutzerkomfort	204
7.4.1	Studie: Social Viewing in CVR.....	204
7.4.2	Studie: Immobile Techniken	211
7.4.3	Studie: Vergleich der Rotationsmedien	215
7.4.4	Studie: Weniger als 360°.....	218
7.4.5	Studie: Vergleich dynamischer und statischer Untertitel	223
7.4.6	Studie: Sprecheridentifikation für Gebärdensprache	230
7.4.7	Studie: Sprecheridentifikation für Untertitel	232
7.4.8	Studie: Kombination world- und screen-referenzierter Untertitel.....	233
7.5	Fazit	236
	Schlusswort	237
	Anhang	238
	Begriffe	238
	Literaturverzeichnis.....	244
	Webseitenverzeichnis	264
	Filmverzeichnis.....	268

Vorwort

Die meisten Menschen, die zum ersten Mal einen Film über ein Head-Mounted Display (HMD) sehen, sind fasziniert von der neuen Erlebniswelt. Das Gefühl, an einem anderen Ort zu sein, mitten im Geschehen, weit weg von der Realität, beeindruckt und lässt sie in eine andere Welt eintauchen. YouTube, Facebook, Magazine (z. B. National Geographic), Zeitungen (z. B. New York Times) und Fernsehkanäle ergänzen ihre Inhalte mit diesen Filmen. Engagierte Filme wie „Clouds Over Sidra“ (Arora und Milk, 2015) oder „Protectors“ (Bigelow, 2017) lassen den Betrachter die Welt der Protagonisten miterleben.

In *Cinematic Virtual Reality* (CVR) taucht der Betrachter in eine mit omnidirektionalen Kameras aufgenommene Umgebung ein. Dabei werden Virtual Reality (VR) Geräte wie HMDs oder Cardboards genutzt. Um mit diesem Medium Geschichten zu erzählen oder Wissen zu vermitteln, sind neue Herangehensweisen erforderlich. Auch wenn CVR-Applikationen in mancher Hinsicht traditionellen Filmen sehr ähneln, die über Jahrzehnte entwickelte Filmsprache lässt sich nicht ohne Weiteres übertragen. In CVR hat der Filmemacher nicht mehr die volle Kontrolle über die vom Zuschauer gesehenen Filmbilder. Manche klassischen Filmtechniken sind nicht anwendbar, aber die dazugewonnene Raumkomponente bietet neue Optionen für intuitive Interaktion und eine individuelle Benutzerlenkung. Dadurch wird es möglich, dass jeder seine eigene Geschichte in seinem eigenen Tempo und mit einem individuellen Fokus entdecken kann.

Herausforderungen

Auch wenn CVR von den Erfahrungen der Filmtheorie und VR-Forschung profitieren kann, gibt es verschiedene Aspekte, die neu betrachtet werden müssen.

Aufmerksamkeitslenkung (Guiding)

In CVR kann der Betrachter die Blickrichtung und damit den sichtbaren Ausschnitt des Films frei wählen. Aus diesem Grund ist es dem Filmemacher nicht verlässlich möglich, dem Zuschauer zu zeigen, was für die Geschichte wichtig ist. Einige klassische filmische Methoden zur Aufmerksamkeitslenkung – wie Nahaufnahmen oder Zooms – sind in CVR nicht praktikabel und müssen durch andere Techniken ersetzt werden. Weitere – wie Bewegung und Farben – benötigen einer Evaluation und Anpassung.

Kamerapositionierung

Da in CVR der Zuschauer genau die Position einnimmt, an der die Kamera während der Dreharbeiten platziert war, sind für die Festlegung des geeignetsten Standpunktes andere Aspekte wichtig als in der klassischen Filmgestaltung. Unterschiedliche Abstände zu Charakteren können ähnlich empfunden werden wie in der realen Welt. Darüber hinaus ist immer nur ein vom Zuschauer selbst ausgewählter Ausschnitt der vom Filmemacher gestalteten Rundumszenerie sichtbar, was bei der Kamerapositionierung berücksichtigt werden muss. Es ist notwendig darüber nachzudenken, ob es dem Zuschauer möglich sein sollte, mehrere Objekte bzw. Charaktere gleichzeitig wahrzunehmen, oder ob Kopfbewegungen dazu notwendig sind.

Montage

In klassischen Filmen wird vom Filmemacher auf einer Timeline der genaue Filmablauf festgelegt. Für die Erstellung eines CVR-Erlebnisses steht eine weitere Komponente zur Verfügung: der Raum. Deshalb sollten Schnitte nicht nur von der verstrichenen Zeit abhängen, sondern auch von der Blickrichtung des Betrachters.

Interaktion

CVR ist von Natur aus ein interaktives Medium. Sich umzusehen und dabei den Bildausschnitt zu wählen, ist eine sehr natürliche und intuitive Art der Interaktion, die dem Betrachter die Kontrolle über das Sichtfeld ermöglicht. Für einige Story-Konstruktionen kann es erforderlich sein, dem Benutzer zusätzliche

Kontrolle zu geben und Abläufe bewusst zu steuern. Diese Interaktionsmöglichkeiten sollten ebenso natürlich und intuitiv sein.

Soziale Aspekte

Viele Menschen schauen Filme gern mit anderen gemeinsam an, was in CVR schwieriger zu realisieren ist. Selbst wenn sich Betrachter gleichzeitig in derselben virtuellen und synchronisierten Umgebung befinden, sehen sie nicht automatisch den gleichen Bildausschnitt. Es ist daher schwieriger miteinander zu kommunizieren. Die Betrachter können sich durch die HMDs isoliert fühlen, was das Benutzererlebnis beeinträchtigen kann.

Analysemethoden

CVR ist als Forschungsgebiet noch relativ neu und es bedarf neuer Methoden, um das Verhalten der Nutzer analysieren zu können. Besonders wichtig dabei sind die Kopf- und Augenbewegungen der Zuschauer, die im Zusammenhang mit dem sich ständig änderndem Bild untersucht werden müssen.

Wissenschaftliche Beiträge

Taxonomie für Guiding Methoden

In der vorliegenden Arbeit wird eine Taxonomie für Methoden zur Lenkung der Aufmerksamkeit in CVR vorgestellt. Diese Taxonomie ermöglicht eine Klassifizierung der Methoden und eine Einschätzung, welche Attribute in bestimmten Anwendungsfällen benutzt werden sollten.

Richtlinien für Kamerapositionen und -höhen

Die Ergebnisse unserer Nutzerstudien zur Positionierung der Kamera in CVR wurden in Guidelines umgesetzt und sollen Filmemacher in der Wahl der richtigen Kamerapositionen und -höhen unterstützen.

Spaceline-Konzept

Das in dieser Arbeit vorgestellte Spaceline-Konzept ermöglicht es, Szenenwechsel nicht nur auf einer Timeline zu definieren, sondern die Blickrichtung der Zuschauer zu nutzen, um CVR-Applikationen interaktiver zu gestalten und nichtlineare Storylines zu realisieren. Dabei ergeben sich zusätzliche Möglichkeiten der Aufmerksamkeitslenkung.

Design-Guidelines für CVR Social Viewing Applikationen

In Design-Guidelines für CVR-Applikationen wurden die Ergebnisse mehrerer Nutzerstudien zusammengefasst. Dabei wurde besonders auf die folgenden Aspekte Wert gelegt: Barrierefreiheit, Benutzerkomfort und gemeinsames Erleben von CVR.

Analyse von Tracking-Daten

Um erkennen zu können, ob Methoden die Blicke der Zuschauer lenken, wurde ein Analysetool entwickelt, welches sowohl Visualisierungen der Gesamtdaten für jeden einzelnen Timecode, wie auch die Blickdaten jedes einzelnen Zuschauers anzeigen kann.

Struktur der Arbeit

Nach einer Einleitung werden in den Kapiteln 2 bis 8 die oben vorgestellten Themen erörtert. Die Nutzerstudien, die in diesen Kapiteln diskutiert werden, sind detailliert in Kapitel 9 beschrieben. Dadurch ist es einfacher, die Ergebnisse unter einander in Beziehung zu bringen und von unterschiedlichen Stellen aus auf sie Bezug zu nehmen. Außerdem ist es dadurch möglich die Analysen und Konzepte im Hauptteil kompakter darzustellen, ohne die notwendige detaillierte Darstellung der Studien zu reduzieren.

Kapitel 1: Einleitung

Das erste Kapitel gibt eine Einführung in das Thema. Ausgehend von einem kurzen historischen Abriss der Film- und VR-Geschichte, werden Begriffe erklärt und Unterschiede herausgearbeitet. Zur Einordnung des Forschungsthemas werden gesellschaftliche Aspekte zum Einsatz von Cinematic VR diskutiert.

Kapitel 2: How to Guide - Aufmerksamkeitslenkung in CVR

Um eine einheitliche Taxonomie für Methoden der Aufmerksamkeitslenkung zu finden, werden Techniken aus Film, Virtual und Augmented Reality analysiert. Mithilfe der erarbeiteten Taxonomie werden Methoden für CVR-Applikationen analysiert und die folgenden Fragen beantwortet:

- Welche Methoden zur Lenkung der Aufmerksamkeit werden in verschiedenen Medien genutzt?
- Durch welche Eigenschaften zeichnen sich diese Methoden aus und wie lassen sie sich klassifizieren?
- Welche Attribute in der Taxonomie sind für welche CVR-Szenarien geeignet?

Kapitel 3: How to Shoot - Kameraführung in CVR

Das dritte Kapitel beschäftigt sich mit der Frage, wie eine CVR-Kamera im Raum platziert werden soll. Dabei werden sowohl Kamerahöhen wie auch Kameraabstände untersucht. Die gewonnenen Erkenntnisse sind in Richtlinien für Filmemacher zusammengefasst. Es werden die folgenden Fragen beantwortet:

- Wie wirken sich Kamerahöhen aus, die nicht der Augenhöhe des Zuschauers entsprechen?
- Was bewirken unterschiedliche Kameraabstände in Bezug auf die Beziehung zu den Charakteren?
- Welche Kamerapositionen sind für den Zuschauer angenehm?

Kapitel 4: How to Cut - Montage in CVR mittels Spaceline-Konzept

Dieses Kapitel untersucht die erweiterten Möglichkeiten der Postproduktion in CVR, wo sich einige Aktionen vom Filmemacher zum Zuschauer verlagern. Die Tatsache, dass der Betrachter das Bild frei wählen kann, bringt neue Herausforderungen aber auch neue Lösungsansätze mit sich. Dieses Kapitel beschreibt das von uns entwickelte Spaceline-Konzept. Verschiedene Montagetechniken, Indikator-Visualisierungen und Selektionstechniken werden analysiert und die Ergebnisse in einem Design-Space und Anforderungen für ein CVR-Editor-Tool zusammengefasst. Es werden die folgenden Fragen beantwortet:

- Welche neuen Begriffe werden für CVR benötigt und wie können diese definiert werden?
- Was bedeuten klassische Filmregeln wie Achsensprung- und 30°-Regel in CVR?
- Welche Montagetechniken sind geeignet, um die beschriebenen Herausforderungen zu meistern?
- Welche Interaktionstechniken werden in verschiedenen Medien genutzt?
- Durch welche Eigenschaften zeichnen sich diese Techniken aus und wie lassen sie sich klassifizieren?

Kapitel 5: How to Watch - Nutzererlebnis in CVR

Kapitel 6 beschäftigt sich mit Aspekten des Nutzerkomforts. Es wird untersucht, wie eine CVR-Applikation gemeinsam erlebt werden kann, und wie auch in bequemen Sitzpositionen ohne zusätzliche Körperrotation ein immersives Erlebnis möglich ist. Des Weiteren werden Aspekte der Barrierefreiheit betrachtet und die Ergebnisse in Design-Guidelines für CVR-Anwendungen zusammengefasst. Es werden die folgenden Fragen beantwortet:

- Welche Techniken können verwendet werden, CVR-Applikationen gemeinsam zu erleben?
- Welche Techniken können genutzt werden, um CVR-Applikationen in bequemen Sitzpositionen – ohne zusätzliche Körperdrehungen – zu erleben?
- Welche Möglichkeiten gibt es, CVR-Applikationen barrierefrei zu gestalten?

Kapitel 6: How to Evaluate – Analysemethoden für CVR

In diesem Kapitel werden verschiedenen Analysemethoden erläutert, die in den Nutzerstudien dieser Arbeit benutzt wurden. Um die Blickrichtung des Betrachters zu untersuchen, wurde ein Tool – der CVR-Analyzer – entwickelt, welches es ermöglicht, Head- und Eyetracking-Daten synchron zum verwendeten Film zu visualisieren.

Kapitel 7: Nutzerstudien

Die Nutzerstudien sind in einem separaten Kapitel am Ende der Arbeit zusammengefasst. Die Studien und ihre Ergebnisse werden hier beschrieben, die Diskussion und Interpretation der Ergebnisse sind in den themenbezogenen Kapiteln zu finden.

Kapitel 8: Zusammenfassung und Ausblick

Das letzte Kapitel widmet sich der Zusammenfassung der wichtigsten Ergebnisse und einem Ausblick auf die weitere Forschung in Cinematic VR.

Stilistische Entscheidungen

Personen: Beim Schreiben über Personen, umfasst das grammatikalische maskuline Geschlecht, z. B. „der Nutzer“ alle biologischen Geschlechter.

Plural: Wie in der HCI-Literatur üblich, wird „wir“ anstelle von „ich“ verwendet, um die Arbeit aller beteiligten Personen und Diskussionspartner zu würdigen.

Filmzitate: Filme werden direkt im Text mit Autor, Titel und Erscheinungsjahr zitiert. Sie erscheinen in einem gemeinsamen Film- und Web-Seitenverzeichnis am Ende der Arbeit. Falls der Film online verfügbar ist, wird dort der Link angegeben.

Übersetzungen: Um Begriffe der englischsprachigen, wissenschaftlichen Literatur besser zuordnen zu können, werden Fachbegriffe in der Regel nicht ins Deutsche übersetzt. Ausnahmen bilden Begriffe, die in der deutschen Sprache schon eine etablierte Übersetzung haben. In diesen Fällen wird die englische Übersetzung bei der ersten Verwendung aufgeführt. Fragen in Nutzerstudien, die aus englischsprachigen Fragebögen stammen, werden nicht übersetzt, um Fehlinterpretationen zu vermeiden.

Zahlen: Aufgrund der besseren Lesbarkeit und der Verwendung englischsprachiger Tools in der Auswertungsphase wird der Punkt statt des Kommas als Dezimaltrennzeichen genutzt.

Begriffe: In der Regel werden Begriffe bei ihrer ersten Verwendung erklärt. Allgemeine Begriffe der Filmtheorie werden in den Kapiteln erläutert, in denen näher auf ihre Bedeutung eingegangen wird. Um das Lesen zu erleichtern werden alle Begriffserklärungen im Anhang zusammengefasst.

Viele Begriffe werden in verschiedenen Fachgebieten in unterschiedlicher Weise verwendet und es kann dabei zu Begriffskonflikten kommen. Im Bereich Mensch-Maschine-Interaktion werden Erkenntnisse aus mehreren Bereichen genutzt, insbesondere der Informatik und der Psychologie. Einige der aus der Psychologie übernommenen Begriffe haben in der Informatik eine andere Bedeutung. In der Psychologie beschreibt der Term *Methode*, das wissenschaftliche Vorgehen, um neue Erkenntnisse zu gewinnen. *Methoden* in der Informatik sind Prozeduren oder Funktionen, die das Verhalten von Objekten beschreiben. Die Benutzung des Terms *Methode* in dieser Arbeit orientiert sich dabei an letzterem: Es werden z.B. Methoden für Szenenwechsel und Interaktionen implementiert und miteinander verglichen.

Für eine leichtere Lesbarkeit und bessere Übersicht werden die folgenden Umrahmungen genutzt:

Beschreibung der in den Nutzerstudien untersuchten Methoden

Zusammenfassung von Ergebnissen

Abkürzungen

AR	Augmented Reality
CAVE	Cave Automatic Virtual Environment
CVR	Cinematic Virtual Reality
DoF	Degree of Freedom
FoV	Field of View
HMD	Head-Mounted Display
IPQ	Igroup Presence Questionnaire
MW	Mittelwert
PiP	Picture-in-Picture
PQ	Presence Questionnaire
RoI	Region of Interest
SD	Standard Deviation
SGD	Subtil Gaze Direction
SMT	Saliency Modulation Technique
SSQ	Simulator Sickness Questionnaire
UEQ	User Experience Questionnaire
VR	Virtual Reality

1

Einleitung

„Die Filmsprache wurde in dem Augenblick geboren, als Filmmacher erstmals den Unterschied erkannten zwischen dem beliebigen Aneinanderreihen von Bildern in unterschiedlichen Stadien der Bewegung und der Anordnung solcher Bildsequenzen in der Absicht, Bezüge zwischen ihnen herzustellen. Sie fanden heraus, dass sie durch die Kombination zweier verschiedener Bildinhalte eine ganz neue Bedeutung erzeugen und Gefühle, Ideen, Tatsachen auf eine Weise mitteilen konnten nach der auch innerhalb anderer Kommunikationssysteme gültigen Formel, dass eins plus eins manchmal drei ergibt. Theoretische Überlegungen führten zu Experimenten.“

Daniel Arijon (Arijon, 2000)

1.1 Historischer Rückblick

1.1.1 Vom bewegten Bild zum Film

Schaut man zurück auf die Anfänge der Filmgeschichte, findet man viele Parallelen zu den heutigen Entwicklungen in Cinematic VR. In den ersten Jahren genügte allein die Attraktion der bewegten Bilder, um die Zuschauer zu begeistern. Die Filme waren nur wenige Sekunden lang und folgten noch keiner Handlung (Roundhay Garden Scene, Louis Le Prince, 1888) (Abbildung 1-1, links). Anfangs ähnelten die Filme lebendigen Fotografien, berühmte Beispiele dafür sind (La Sortie De l'Usine Lumière À Lyon, Louis Lumière, 1895) und (Arrival of a Train at La Ciotat, Lumière Brothers, 1895).



Abbildung 1-1: links: Roundhay Garden Scene (Louis Le Prince, 1888), rechts: Le Voyage dans la lune (Méliès, 1902).

Kurze Zeit später wurden schon erste Geschichten erzählt (Le Voyage dans la lune, Georges Méliès, 1902) (Abbildung 1-1, rechts), dabei nahm die Kamera anfangs stets aus dem gleichen Blickwinkel und mit der gleichen Einstellungsgröße auf – so wie man es aus dem Theater gewohnt war. Allerdings fiel ein Mittel der Aufmerksamkeitslenkung, welches im Theater zur Verfügung stand, in der Stummfilmzeit weg. Mit Geräuschen und Stimmen konnten im Theater Zuschauer in eine Bühnenregion gelenkt werden.

Anfang des 20. Jahrhunderts wurde zum ersten Mal eine Nahaufnahme verwendet (The Sick Kitten, G.A. Smith, 1903) (Abbildung 1-2, links). Mit den Jahren entwickelte sich eine eigene Erzählkunst mit Perspektivenwechseln, verschiedenen Einstellungsgrößen und unterschiedlichen Rhythmen im Schnitt. Was aus heutiger Sicht selbstverständlich scheint, musste erst entdeckt werden. Der Zauberkünstler und Filmemacher Georges Méliès experimentierte mit unterschiedlichen Schnitteffekten, wie Fade-In und Fade-Out, Überblendungen und Stop-Motion und wendete sie erfolgreich in seinen Filmen an (Long Distance Wireless Photography, Georges Méliès, 1908) (Abbildung 1-2, rechts). Schon bald wurden erste Versuche unternommen, dieses neue Medium Film von der Theaterperspektive zu lösen und als etwas Neues zu verstehen, weder als eine Erweiterung der Fotografie mit bewegten Bildern, noch des Theaters, mit Aufnahmen vom besten Zuschauerplatz. Es wurden verschiedene Einstellungsgrößen miteinander verbunden und Aufnahmen verschiedener Orte miteinander kombiniert.

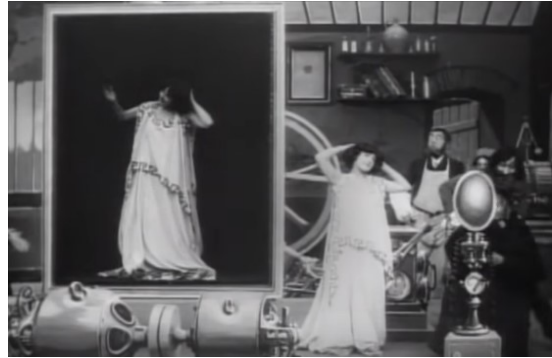


Abbildung 1-2: links: *The Sick Kitten* (Smith, 1903), rechts: *Long Distance Wireless Photography* (Méliès, 1908).

Die ersten Filme mit verschiedenen Handlungssträngen entstanden. Die Zeit zwischen zwei Schnitten war damals wesentlich länger und wirkt aus unserer heutigen Sicht langatmig. Andere damals übliche Erzählweisen muten mit unseren heutigen Sehgewohnheiten ungewohnt an. Noch wurde zwischen zwei Szenen an unterschiedlichen Orten ab- und aufgeblendet, da man glaubte, der Zuschauer könnte sonst den Ortswechsel nicht realisieren. Im Jahre 1903 zeigte Porter einen Film, in dem Parallelmontage und harte Schnitte verwendet wurden (*The Great Train Robbery*, Porter, 1903). Schnitte gab es ab diesem Zeitpunkt nicht nur am Ende einer Szene, sondern auch mittendrin. Nun war nicht länger die Szene die Basiseinheit des Films, sondern die Einstellung.

Aber schon bald ließ der Reiz des neuen Mediums nach und die bisherigen Inhalte genügten nicht mehr, um die Zuschauer zu begeistern. Neue Wege für das Erzählen von Geschichten durch Filme wurden gesucht. Die Filmsprache entwickelte sich weiter, Filmemacher und Zuschauer lernten mit neuen Methoden umzugehen und ihre Sehgewohnheiten anzupassen. Auch die Entwicklung neuer Technik ermöglichte neue Elemente der Filmsprache. Im Jahre 1914 wird das erste Mal eine Kamerafahrt in einem Film gezeigt (*Cabiria*, Giovanni Pastrone, 1914).

Im Laufe der Jahre hat sich eine eigene Erzählkunst entwickelt, die auch Gegenstand wissenschaftlicher Untersuchungen wurde. Die narrativen Methoden der traditionellen Filmproduktion lassen sich jedoch nicht ohne Weiteres auf CVR übertragen. Der Wechsel einiger Aktionen vom Filmemacher zum Zuschauer und neue Interaktionsmöglichkeiten erfordern und ermöglichen neue Ansätze.

1.1.2 Panorama-Filme, 360°-Filme und Sphärische Filme

Schon seit Jahrhunderten erschaffen Menschen Bilder nicht nur als flache Abbildungen, sondern erzeugen durch Rundumanordnungen räumliche Erfahrungen: angefangen bei Höhlenmalereien über Triptychen bis zu den ersten 360°-Gemälden in extra dafür errichteten Gebäuden im 18. Jahrhundert. Das erste dieser Gebäude zeigte das im Jahre 1787 von Robert Barker erschaffene Panorama von London (Abbildung 1-3, links). Barker selbst verband sein Gemälde mit dem Gefühl vor Ort zu sein (Ford, 2017). Mit der Entwicklung der Fotografie wuchs auch der Wunsch, Panoramabilder in dieser Technologie zu erzeugen. 1843 erfand der Österreicher Joseph Puchberger die erste Panoramakamera, die allerdings erst 150° aufnehmen konnte (Abbildung 1-3, Mitte). Im Jahre 1887 wurde die erste Kamera patentiert, die 360° auf einer Filmrolle aufzeichnen konnte (Abbildung 1-3, rechts) (Lomography, 2017).

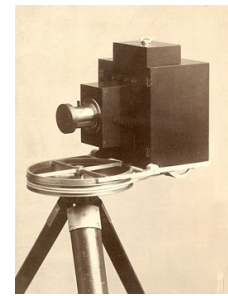
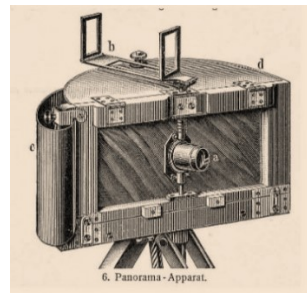
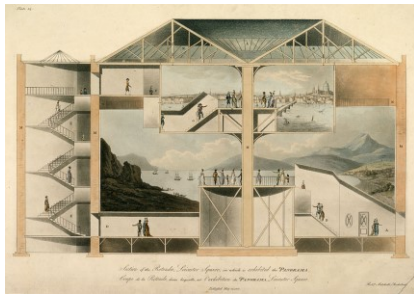


Abbildung 1-3: links: Rotunda am Leicester Square in London (British Library, 2017), Mitte: erste Panoramakamera 1843 (media+art+innovation, 2014), rechts: erste 360°-Kamera 1887 (Lomography, 2017).

Auch sphärische Filme haben bereits eine lange Tradition. Die erste 360°-Leinwand wurde im Jahre 1900 bei der Weltausstellung in Paris vorgestellt (Lescop u. a., 2017). Sie hatte einen Umfang von 100 m. Mittels 10 Projektoren wurden statische gemalte Bilder einer Ballonfahrt auf die Panoramaleinwand projiziert, um damit dem Zuschauer, der in einer Zuschauerkabine in Form einer Ballongondel stand, das Gefühl zu vermitteln, selbst im Ballon zu sein (Abbildung 1-4, links). Die Projektoren in der Kabine liefen schnell heiß. Es gibt verschiedene Versionen darüber, ob die Premiere wirklich stattfand (Piccolin, 2006). Aber die Idee, den Menschen mit der Filmszenerie zu umgeben, lebte weiter fort.

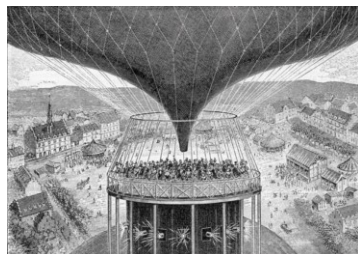


Abbildung 1-4: Links Cinéorama Projektion auf der Weltausstellung von Paris (1900)(Wikipedia, 2020), rechts Circarama auf der Expo in Lausanne (1964)(SRF, 2018).

Als im Jahre 1955 in Kalifornien der erste Disney-Park eröffnete, war ein 360°-Kino mit 12 m Durchmesser eine der Attraktionen (Piccolin, 2012). Die ersten Kamerasysteme dieser Zeit brachten noch das Problem mit sich, dass die elf 16-mm-Kameras so angebracht werden mussten, dass zwischen den Aufnahmen Teile der 360°-Szene fehlten, die während der Projektion durch vertikale Bildschirmteiler sichtbar waren (Abbildung 1-5, links, Abbildung 1-4, rechts). Der damalige Disney-Techniker Ub Iwerks fand eine Lösung: Er ordnete die Kameras so an, dass sie nach oben in Spiegel filmten, die in einem Winkel von 45° angeordnet waren. Der Abstand vom Schnittpunkt des Spiegels zur Linse war der gleiche wie der Abstand von diesem Punkt zur Mitte des Systems. Dadurch filmten die Kameras so, als ob alle Linsen genau in der Mitte wären (Abbildung 1-5, rechts) (Iwerks, 2019).

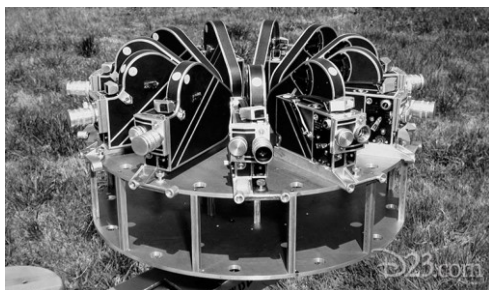


Abbildung 1-5: 360°-Kamerasysteme der 1950er und 1960er-Jahre. (SRF, 2018).

1.1.3 VR und AR

Virtual Reality ist kein neues Medium. Bereits in den 1960er-Jahren wurden Geräte entwickelt, die die Grundlagen für heutige VR-Anwendungen schufen (Jerald, 2015). Eines der frühesten bekannten Beispiele für immersive, multisensorische Technologie, das verschiedene Sinne ansprach, ist das Sensorama, welches 1962 vom Kameramann Morton Heilig entwickelt wurde (Abbildung 1-6, links). Das System ermöglichte Stereoton, war in der Lage zu vibrieren, Gerüche zu erzeugen und neben Hitze auch Wind zu generieren. Der Zuschauer konnte eine 10-minütige Motorradfahrt durch New York als multisensorielle Erfahrung erleben. Das Sensorama gilt heute als ein Vorläufer von Virtual-Reality-Systemen.

In den 1960er-Jahren entwickelte Ivan Sutherland das erste HMD mit Motion Tracking (Abbildung 1-6 ,rechts). Der Apparat war so schwer, dass er an der Decke befestigt werden musste, damit der Träger von dessen Gewicht nicht erschlagen wurde, so dass seine Entwickler ihn auch liebevoll das Damoklesschwert nannten.

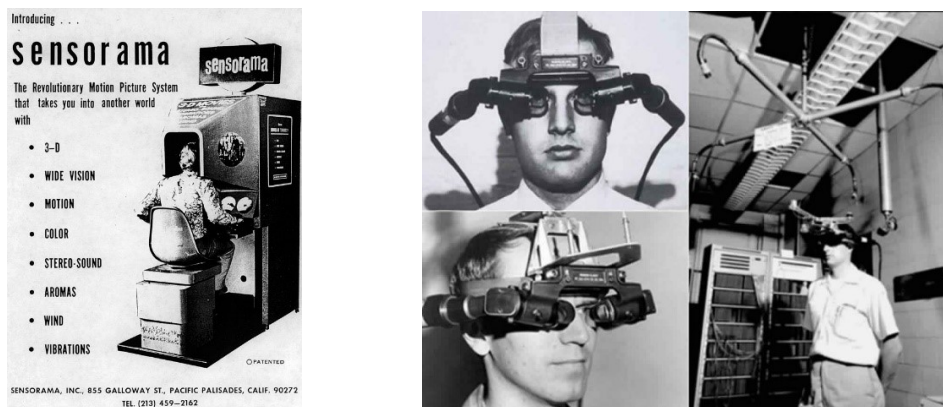


Abbildung 1-6: links: Sensorama (1962) (Reismann, 2020), rechts: Sutherlands Damoklesschwert (1965/1968) (Kaur, 2018).

Es dauerte jedoch einige Zeit, bis die Öffentlichkeit Zugang zu VR-Geräten hatte. Der Begriff *Virtual Reality* wurde das erste Mal von Jaron Lanier, einem amerikanischen Informatiker und Musiker verwendet. Er war Mitbegründer von VPL Research, dem ersten kommerziellen Virtual Reality Unternehmen, welches ein HMD (EyePhone) und einen Datenhandschuh auf den Markt brachte (Point in Time Studios, 2018; Swenberg und Eriksson, 2018). In den 1990er-Jahren wurden Spiele für HMDs entwickelt, aber der Komfort der Geräte und der technische Fortschritt, insbesondere im grafischen Bereich, reichten nicht aus, um die Masse zu erreichen. Durch die technischen Entwicklungen der letzten Jahre wurden die Geräte kleiner, billiger und komfortabler.

1.1.4 Cinematic VR

All die beschriebenen historischen Entwicklungen legten den Grundstein dafür, dass sich Cinematic VR entfalten konnte. Sowohl VR als auch Cinematic VR erlebten ab 2012 eine rapide Entwicklung: 360° Linsen für kleine Geräte kamen auf den Markt (Byford, 2012), Oculus startete eine erfolgreiche Kickstarter-Kampagne und wurde 2014 von Facebook für 2 Milliarden US-Dollar aufgekauft, was viel Aufsehen erregte und einen Boom an neuen VR-Geräten auslöste. Zahlreiche HMDs und Kameras, die für Konsumenten erschwinglich waren, eroberten den Markt, und seit 2014 unterstützen Facebook und YouTube 360° Videos (sphärische Videos).

Das Sundance Filmfestival im Jahre 2012 war das erste Filmfestival, auf dem VR-Filme gezeigt wurden (Dickey, 2014; Yabus, 1967). Neue sind seitdem entstanden und immer mehr etablierte Festivals haben einen VR Teil im Programm, unter anderem das Tribeca Filmfestival, das Venice Filmfestival und die IDFA (International Documentary Film Festival Amsterdam).

1.2 Was ist Cinematic VR?

1.2.1 Begriffe

Der Begriff Virtuelle Realität wird nicht einheitlich angewandt. In dieser Arbeit verwenden wir die folgende Definition: *Virtual Reality* (VR) ist eine Computertechnologie, die mittels Software Bilder, Töne und andere Sinneseindrücke erzeugt, um einen anderen Ort zu schaffen, so dass der Benutzer das Gefühl hat, Teil dieses anderen Ortes zu sein (Bendel, 2018). Dabei wird das Gefühl der *Presence* ausgelöst, das Gefühl, sich an einem bestimmten Ort zu befinden (Slater, 2003). Während *Presence* eine Reaktion auf eine virtuelle Umgebung ist, beschreibt der Begriff *Immersion* das, was die Technologie aus objektiver Sicht liefert (Slater, 2003). Den Begriff der *immersiven Erfahrung* findet man auch außerhalb der VR-Terminologie. Auch ein Hörspiel oder ein Bild kann als immersive Erfahrung bezeichnet werden. Im hier dargelegten Zusammenhang bezieht er sich aber auf das Eintauchen in eine andere Welt mittels VR-Technologien.

Auch sphärische Filme können dieses Gefühl der Immersion und der *Presence* erzeugen, wenn sie durch *Head-Mounted Displays* (HMDs) oder andere VR-Geräte (z. B. CAVE) betrachtet werden. In der vorliegenden Arbeit wird der Begriff *Cinematic VR* (CVR) für dieses Medium verwendet. Dabei nimmt der Betrachter den Film als virtuelle Realität wahr und befindet sich innerhalb der Filmszene.

Der in der Umgangssprache verwendete Ausdruck *360°-Film* beschreibt das Medium nur teilweise. Zum einen sind es nicht nur 360°, in denen sich der Blick des Zuschauers bewegt, auch ein nach oben und unten Schauen ist möglich. In der wissenschaftlichen Literatur findet man die zutreffenderen Begriffe *omnidirektional* oder *sphärisch* für Medien, die in alle Richtungen Bildinformationen liefern. Weitere verwendete Begriffe für sphärische Filme sind: immersive Filme und VR-Filme. Zum anderen ist es möglich, solche Filme auch an einem Desktop durch Bewegung der Maus oder auf einem Smartphone oder Tablet durch Bewegung des Gerätes anzuschauen. In CVR wird der sphärische Film in einer VR-Umgebung betrachtet, meist mittels eines HMDs. Dabei sind die gleichen Untersuchungsmethoden wie in anderen VR-Gebieten relevant, z. B. die Messung von Auswirkungen auf *Presence* und *Simulator-Sickness*.

Sphärische Filme sind nicht neu und werden nicht nur für HMDs produziert. Untersuchungen über sphärische Videos auf einem Desktop (Boonsuk, 2011) liefern wichtige Grundlagen. Im Gegensatz dazu nennen wir klassische Filme, die für eine Kinoleinwand oder einen TV-Bildschirm produziert werden, *framed Filme*. Der wesentliche Unterschied zu sphärischen Filmen ist der begrenzte Bildausschnitt. Beide Medien können sowohl stereoskopisch als auch monoskopisch vorliegen und sind im monoskopischen Fall 2-dimensional. Während in framed Filmen der Filmemacher bestimmt, was im Bildausschnitt zu sehen ist, kann in CVR der Zuschauer durch die Wahl der Blickrichtung das Sichtfeld (engl. *Field of View*, *FoV*) bestimmen. Um den Zuschauer auf bestimmte Regionen, die *RoIs* (*Region of Interest*), aufmerksam zu machen, kann der Filmemacher bestimmte *Cues* (Hinweise) verwenden.

1.2.2 Vergleich von framed Filmen und Cinematic VR

Auch wenn CVR-Applikationen in manchen Aspekten framed Filmen ähneln, gibt es eine Reihe von Unterschieden, die neue Ansätze und eine eigene Filmsprache erfordern.

Im kreativen Prozess des Filmemachens stehen verschiedene Stilmittel zur Verfügung, mit denen Filmemacher die Geschichte gestalten. In beiden Komponenten des Filmemachens, dem Dreh und dem Schnitt, finden sich Elemente, deren Anwendung oder Bedeutung sich für CVR ändert. Für einige von ihnen wechselt die Verantwortung vom Filmemacher zum Zuschauer. Tabelle 1-1 zeigt einige Unterschiede zwischen herkömmlichen Filmen und CVR.

Framed Film	Cinematic VR (CVR)
Filmemacher wählt den Bildausschnitt	Zuschauer wählt den Bildausschnitt

Filmemacher schwenkt die Kamera	Zuschauer dreht den Kopf
Zoom zur Annäherung	Kein Zoom
Zuschauer ist außerhalb der Szene	Zuschauer ist in der Szene
mehrere Einstellungen für eine Szene (z. B. Totale, Nahaufnahme)	oft nur eine Einstellung für eine Szene, vom Betrachter erkundet
Filmemacher bestimmt das Tempo des Films anhand der Länge der Aufnahmen	Der Betrachter bestimmt das Tempo mit, indem er die Blickrichtung ändert
Bewegte Bilder	
Mit Kameras aufgenommen	

Tabelle 1-1: Unterschiede und Gemeinsamkeiten zwischen framed Filmen und CVR.

Die **Kinematografie**, die Aufzeichnung von fotografischen Bewegtbildern, ist die Kunst, die Zusammensetzung einer Szene zu bestimmen, indem Kameras, Lichter, Personen und Objekte an die beste Position gebracht werden, um die Geschichte zu unterstützen (Kirchlechner, 2013). Elemente der Kinematografie sind beispielsweise Kamerabewegungen, Bildausschnitte, Bildgrößen, Szenenbilder, Kameraperspektiven und Brennweiten. Die meisten von ihnen sind auch in CVR verfügbar. Einige können jedoch nicht mehr vom Filmemacher allein bestimmt werden. In CVR wechseln verschiedene Entscheidungen vom Filmemacher zum Zuschauer.

Kamerabewegung: Beim herkömmlichen Filmemachen gibt es zwei Arten von Kamerabewegungen: Ändern der Position einer Kamera und Drehen der Kamera. Solche Bewegungen können für verschiedene Zwecke genutzt werden, z. B. um die Aufmerksamkeit des Betrachters zu lenken, verschiedene Orte zu verbinden oder visuelle Effekte zu erzeugen. Auch wenn eine Fahraufnahme nicht der Bewegung des CVR-Betrachters entspricht, der sich innerhalb der Szene befindet, werden manche Bewegungen vom Betrachter akzeptiert. Die zweite Bewegungsart, das Schwenken der Kamera, ist nicht erforderlich, da den Zuschauern der gesamte Film in allen Richtungen zur Verfügung steht. Eine Drehung der Kameraansicht innerhalb einer Szene, die nicht durch den Betrachter ausgelöst wurde, kann zur Reisekrankheit führen, da die tatsächliche Kopfbewegung des Betrachters nicht mit der Drehung in der virtuellen Welt korrespondiert. Im Gegensatz zur Fortbewegung sind CVR-Betrachter es gewohnt, die Blickrichtung selbst zu ändern.

Framing: Für den klassischen framed Film wählen Filmemacher den Bildausschnitt aus, der für den Betrachter sichtbar ist. Grenzen und Zusammensetzung der Bilder bestimmen die Aussagekraft von Subjekten und Personen. In CVR wird diese wichtige Ressource den Zuschauern anvertraut, sie tun dies, indem sie den Kopf bewegen. Mit einer qualifizierten Kameraposition können Filmemacher jedoch den Inhalt des ausgewählten Bildes beeinflussen. Mit den Abständen zwischen Motiven / Objekten und Kamera können Größen und Abstände bestimmt werden, ähnlich wie bei herkömmlichen Filmen (z. B. Nahaufnahme, Schulter-Nahaufnahme, Totale). Das Ändern des Kameraabstands ist jedoch nur in einer Hauptrichtung möglich. Wenn die Kamera in einer Richtung näher an einem Objekt positioniert ist, bewegt sie sich von einem Objekt auf der gegenüberliegenden Seite weg.

Mis-en-scène (placing on stage): Das Anordnen von Personen und Objekten zueinander und zur Kamera beeinflusst, wie die Objekte / Personen und ihre Verbindungen zueinander wahrgenommen werden. In CVR unterscheidet sich die Zusammensetzung des Raums vom traditionellen Filmemachen. Während ein traditioneller Filmemacher die Szene in einem begrenzten Bereich arrangieren kann, müssen CVR-Filmemacher über den gesamten Raum nachdenken. Zusätzlich kann der traditionelle Filmemacher entscheiden, ob zwei Personen zusammen auf einem Bild zu sehen sind oder nicht. In CVR beeinflusst der Benutzer diese Tatsache durch die eigene Blickrichtung. Der Filmemacher entscheidet jedoch, ob Personen so nah beieinander sind, dass sie gleichzeitig gesehen werden können oder so weit voneinander entfernt sind, dass der Betrachter den Kopf bewegen muss, um den einen oder den anderen zu sehen.

Kamerahöhe: Ähnlich wie bei herkömmlichen Filmen kann die Kamerahöhe für die Komposition von Szenen in CVR verwendet werden. Eine niedrige Kamera schaut hinauf und macht den Charakter in den

meisten Fällen bedeutungsvoller und mächtiger. Eine erhöhte Kamera schaut auf ein Objekt oder einen Charakter herab, der Charakter scheint kleiner zu sein. Dabei muss berücksichtigt werden, dass in CVR der Zuschauer diese Kameraposition in der virtuellen Welt einnimmt. Der Normalfall, die Kamera auf Augenhöhe zu platzieren, ist in Cinematic VR aufgrund der unterschiedlichen Körpergrößen nicht so einfach zu realisieren.

Filmschnitt/Montage ist der Prozess der Auswahl und Anordnung von Aufnahmen für die Produktion des Films. Die Aufnahmen werden zu einer Sequenz zusammengefügt und daraus ein Film. Elemente der Bearbeitung sind beispielsweise Schnitte, Übergänge, Rhythmus und Tempo.

Schnitte: Schnitte haben beim Filmemachen unterschiedliche Funktionen: Sie verkürzen die Dauer eines aufgezeichneten Vorgangs, machen Ortswechsel möglich, schaffen Nähe oder Distanz. Die Länge eines Schnitts beeinflusst das Tempo und den Rhythmus des Films. Das perfekte Verbinden von Aufnahmen ist eine Kunst, die den Film erheblich beeinflusst. Auch wenn wir in der realen Welt keine Schnitte und Übergänge haben, akzeptieren die Zuschauer dies in der virtuellen Welt (Kjær u. a., 2017). Ähnlich wie bei framed Filmen können Fehler beim Verbinden der Aufnahmen den Betrachter irritieren.

Übergänge (Transitions): Um zwei Aufnahmen zu verbinden, werden in framed Filmen verschiedene Übergangstechniken verwendet, z. B. Schwarz- und Weißblenden, Überblendungen und andere Effekte. Einige von ihnen wurden für real-time VR untersucht (Medeiros u. a., 2016; Men u. a., 2017; Oberdörfer u. a., 2018). Es bedarf weiterer Nachforschungen zu solchen Übergängen in CVR.

Rhythmus: Der Rhythmus eines Films wird von Tempo und Timing beeinflusst. Lange Einstellungen erzeugen ein Gefühl der Langsamkeit, kurze ein Gefühl der Schnelle. In framed Filmen bestimmt allein der Filmemacher den Rhythmus durch das Zusammensetzen der Einstellungen in der Timeline. Einstellungen in CVR müssen länger sein als in framed Filmen, da der Betrachter mehr Zeit zum Umsehen benötigt. Wie schnell dies getan wird, beeinflusst auch den Rhythmus des Filmelerlebnisses. Der Filmemacher kann dies durch die Komposition des Bildes beeinflussen, insbesondere durch die Anzahl der Objekte, die verschiedenen Farben und die Verschiedenartigkeit der Bilder.

1.2.3 Vergleich zwischen Game und Cinematic VR

Auch wenn Stories in CVR in der Regel noch immer linear sind, ist dieses Medium sehr gut geeignet, um interaktive nichtlineare Stories zu erschaffen. Dafür übernehmen wir einige Begriffe aus der Game-Theorie. Sylvester (Sylvester, 2013) beschreibt drei Typen von Stories: *scripted Stories*, *World Narratives* und *emergent Stories*.

Scripted Stories laufen immer auf die gleiche Weise ab und können vom Nutzer nicht beeinflusst werden. Viele CVR-Anwendungen sind momentan *soft-scripted* Stories. Eine *soft-scripted* Story läuft zwar immer gleich ab, aber ohne die Benutzeroberfläche des Nutzers zu deaktivieren (Sylvester, 2013). Der CVR-Nutzer kann sich mittels HMD umsehen und den Bildausschnitt steuern.

Eine *World Narrative* ist die Geschichte eines Ortes, seiner Vergangenheit und seiner Menschen. Sie wird durch die Konstruktion des Ortes und der darin befindlichen Objekte erzählt. Diese Geschichten müssen nicht linear sein. Es ist nicht wichtig, in welcher Reihenfolge die Details entdeckt werden. World Narratives sind für CVR gut geeignet und benötigen keine zusätzliche Aufmerksamkeitslenkung.

Eine *emergent Story* wird durch die Interaktion der Nutzer generiert. Ein Beispiel ist ein Racing-Game, in dem die Story durch die Aktionen der Spieler entsteht. Diese Art von Stories sind für CVR weniger relevant.

Auch wenn die Grenzen zwischen interaktiver CVR und Game fließend sind, unterscheiden sie sich doch in der Art ihrer Konstruktion und dem Ziel der Applikation. Während Games Konstruktionen von Prozessen und Regeln sind, ist CVR eine Konstruktion von Worten und Bildern (Sylvester, 2013). Ein Gamer möchte einen möglichst hohen Score erreichen oder eine Aufgabe lösen (aktiv), ein CVR-Zuschauer gut unterhalten werden (passiv).

1.2.4 Vergleich zwischen Real-time VR und Cinematic VR

In VR kann der Benutzer interaktiv eine digitale Umgebung erkunden. Er nimmt die virtuelle Welt als eine Art Realität wahr und kann ein Teil davon sein. Durch das Betrachten von sphärischen Filmen in einer virtuellen Umgebung schaut der Betrachter nicht mehr von außen auf den Film, sondern befindet sich innerhalb der Szene. In CVR werden HMDs oder andere VR-Geräte verwendet, um Filme anzusehen. Genau wie in computer-erzeugter virtueller Realität, bei der der Inhalt in Echtzeit entsteht (*real-time VR*), kann sich der Betrachter frei umsehen und den sichtbaren Bereich des Bildes auswählen. Während in real-time VR die Inhalte je nach Benutzeraktion angepasst werden können, werden die Inhalte von CVR vorab von Kameras oder in Animationsstudios produziert. Tabelle 1-2 zeigt einige grundlegende Unterschiede und Gemeinsamkeiten zwischen real-time VR und Cinematic VR.

Real-time VR	Cinematic VR
Bilder entstehen in Echtzeit, sind anpassbar	Bilder wurden vorab aufgenommen
Kontinuierliche Bewegungen in der Welt sind möglich	Ortswechsel durch „Teleportieren“
6 Freiheitsgrade (DoF)	3 Freiheitsgrade (DoF)
Display: HMD, CAVE, VR-Devices	
Gefühl, dort zu sein	

Tabelle 1-2: Unterschiede und Gemeinsamkeiten zwischen real-time VR und Cinematic VR.

Milgram und Kishino (Milgram und Kishino, 1994) beschrieben das Kontinuum zwischen realer und vollständig virtueller Umgebung wie in Abbildung 1-7 zu sehen ist. Im Gegensatz zu VR, wo die reale Welt durch die virtuelle ersetzt ist, ist *Augmented Reality* (AR) die Überlagerung der realen Welt durch digitale Inhalte. Dadurch ist es dem Nutzer möglich, sowohl mit der realen Welt wie auch mit digitalen Elementen zu interagieren.

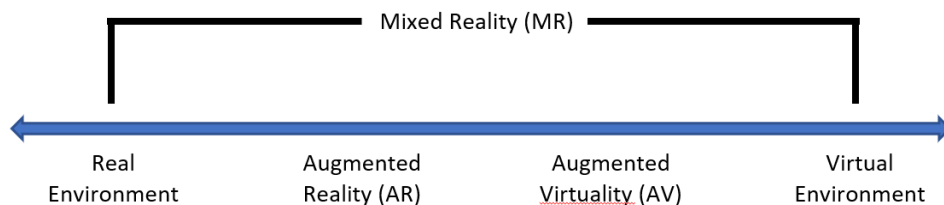


Abbildung 1-7: Darstellung des Virtualitätskontinuums (Milgram und Kishino, 1994).

Trotz aller Unterschiede kann CVR einen Großteil der Erkenntnisse der bisherigen VR-Forschungen nutzen. Themen wie Interaktion benötigen aber neue Ansätze. Die Hauptform der Interaktion – das Umherschauen mit dem HMD und damit das Auswählen des Bildausschnittes ist beiden Gebieten gemeinsam. Die Bewegung im Raum und das Verändern von Objekten der VR-Umgebung ist in CVR jedoch nicht möglich.

1.3 Anwendungsgebiete, Möglichkeiten und Grenzen von CVR

Entertainment

Das Spektrum von Cinematic VR ist breit und reicht wie bei klassischen framed Filmen vom Spielfilm, über den Dokumentarfilm bis hin zu Animationen. CVR-Filme können uns das Weiße Haus von innen zeigen (Felix & Paul Studios, 2017), den Kampf von Rangern im Kongo gegen das Elefantentöten näherbringen (Bigelow, 2017) und die Welt aus der Sicht von Tieren darstellen (iAnimal360, 2016; Marshmallow Laser

Feast, 2015). Die Meinungen darüber, ob sich dadurch ein tieferes Verständnis und bewusstere Verhaltensweisen erzeugen lassen, gehen auseinander (Suh u. a., 2018; Wang u. a., 2018). Der Wunsch in einem Überangebot an Medieninhalten durch neue Techniken mehr Aufmerksamkeit zu erzeugen ist naheliegend, kann aber allein über die Technik nur kurzfristig gelingen. Um den Zuschauern Inhalte anzubieten, die diesem Medium gerecht werden, sind neue Ideen und weitere Forschungen notwendig.

Virtuelles Reisen

Natürlich kann Cinematic VR ein echtes Reiseerlebnis nicht ersetzen. Aber es ist eine Möglichkeit, sich Plätze anzuschauen, in Fällen, in denen Reisen zu aufwendig oder nicht möglich wären. Eine Schulklasse kann innerhalb kurzer Zeit in einem Tagebau stehen, antike Bauwerke in Originalgröße erfahren und die Lebenswelt tropischer Tiere beobachten. Auch von Orten, die aus unterschiedlichen Gründen nicht besucht werden können oder sollten (z. B. isolierte oder bedrohte Regionen), kann man sich so einen Eindruck schaffen – in der Regel ohne schlechtes Wetter und große Menschenmassen. Außerdem kann Menschen, die aus unterschiedlichen Gründen nicht reisen können oder möchten, auf diese Weise ein Eindruck von Plätzen und Umgebungen vermittelt werden.

Bildung

VR und CVR sind interaktive Medien, und Interaktion ist förderlich für Lernprozesse (Evans und Gibbons, 2007). Noch stehen die Forschungen über den Einsatz dieser Medien im Unterricht noch am Anfang (Howard u. a., 2018; Martín-Gutiérrez u. a., 2017; Penn P. Wu, 2017; Roussou, 2004; Stojšić u. a., 2019). Herausforderungen, Möglichkeiten und Grenzen müssen ausgelotet werden. Mittels VR und CVR sind neue Formen experimentellen Lernens möglich: physikalische Gesetzmäßigkeiten können erfahrbar gemacht oder außer Kraft gesetzt, mathematische Grundbegriffe visualisiert und Geschichte erlebbar gemacht werden. Der Einsatz von CVR im Unterricht kann Engagement und Aufmerksamkeit erhöhen (Davies u. a., 2019; Harrington u. a., 2018). Allerdings sind oft keine Vorteile bezüglich der Merkfähigkeit gegenüber konventionellen Lehrmethoden nachweisbar (Harrington u. a., 2018; McKenzie u. a., 2019).

Immersiver Journalismus

Auch der Journalismus hat VR und CVR als Vermittlungsmedium entdeckt. Einige computerbasierte VR-Applikationen lassen den Nutzer an anderen Welten teilhaben (de la Peña, 2014, 2012, 2007) wobei Umgebungen und Charaktere am Computer erzeugte 3D-Objekte sind. Auch wenn es bei diesen Projekten einen wirklichen 3D-Raum mit 6DoF gibt, ist nicht erweisen, ob diese zusätzliche Immersionskomponente mehr Presence und Glaubwürdigkeit erzeugt, als abgebildete reale Umgebungen und Charaktere in CVR.

Medienkonzerne wie die New York Time und die BBC nutzen CVR, um ihr Nachrichtenangebot zu erweitern und mit neuen Technologien attraktiver zu gestalten (New York Times, 2015). Inwieweit VR und CVR Journalismus erweitern kann, welche Herausforderungen und Grenzen es dabei gibt, wird derzeit diskutiert und erforscht (Medieninnovationszentrum Babelsberg, 2018).

Ist VR eine Empathie-Maschine?

Chris Milk erweckte 2015 viel Aufsehen, als er den Begriff der „Empathy Machine“ für Virtual Reality (Milk, 2015) nutzte. Die Diskussion darüber dauert bis heute an und führte zu Auseinandersetzungen über Empathie und ihre Auswirkungen. Der Gedanke mit diesem Medium mehr Empathie auslösen und damit das Verhalten der Menschen beeinflussen zu können, brachte eine Reihe von VR-Projekten und CVR-Filmen hervor, die den Zuschauer die Welt von Obdachlosen (Asher u. a., 2018; Herrera u. a., 2018; Stanford Studie, 2017) oder das Leben in Flüchtlingslagern (Arora und Milk, 2015) nahebringen soll.

Der Psychologe Paul Bloom widerspricht dieser Meinung und kritisiert den Versuch, Mitgefühl für Flüchtlinge mittels VR auslösen zu wollen (Bloom, 2017b). VR gibt nicht das wirkliche Gefühl wieder – nur die Erscheinung der Welt vor Ort, nicht die Angst der Protagonisten, nicht ihre bisherigen Erlebnisse und die unsichere Zukunft. Der Zuschauer kann jederzeit das HMD absetzen und die Anwendung beenden (Bloom, 2017b). Bloom geht noch einen Schritt weiter: Empathie verleitet zu unfairen irrationalen Entscheidungen und sollte nicht als Richtlinie für moralisches Handeln dienen (Bloom, 2017a, 2016).

Die Forschungsergebnisse zu diesem Thema sind unterschiedlich. Eine Ursache dafür ist, dass der Begriff Empathie nicht einheitlich verwendet wird (Soler-Adillon und Sora, 2018). In der Studie von Herrera et al. (Herrera u. a., 2018) unterzeichneten mehr Nutzer einer VR-Applikation über Obdachlosigkeit eine Petition für bezahlbaren Wohnraum für Obdachlose im Vergleich zu denen, die sich über andere Methoden mit dem Thema auseinandersetzten. Die Autoren räumen ein, dass es weiterer Forschungen bedarf, ob der Novelty-Effekt Auswirkungen auf dieses Ergebnis hatte.

CVR-Filme sind erst seit wenigen Jahren im Blick der Öffentlichkeit. Eine Diskussion über Fragen der Ethik und Grenzen von VR und CVR ist notwendig und hat bereits begonnen (Kent, 2015; Slater und Sanchez-Vives, 2016; Steinicke, 2018) und zu ersten Ansätzen für Regeln im Umgang mit VR geführt (Fisher, 2017; Madary und Metzinger, 2016).

1.4 CVR als Forschungsgebiet

CVR ist als Forschungsgebiet noch recht neu, kann aber Erkenntnisse aus Filmtheorie, VR und AR nutzen, modifizieren und weiterentwickeln. Da im Gegensatz zu anderen Gebieten noch keine einheitliche Terminologie existiert, ist ein Ziel dieser Arbeit, vorhandene Literatur zu analysieren, die Begriffe zu klassifizieren und Konzeptideen zu entwickeln.

Wie in anderen HCI-Gebieten werden auch in CVR Nutzerstudien verwendet, um neue Erkenntnisse zu gewinnen. Dabei geht es im gegenwärtigen Stadium vor allem darum, Erfahrungen zu sammeln, um neue Methoden entwickeln zu können. Noch existieren keine evaluierten Fragebögen für CVR. VR-Fragebögen (Kennedy u. a., 1993; Schubert u. a., 2002; Witmer und Singer, 1998) können nur bedingt eingesetzt werden, helfen aber beim Sammeln von Erfahrungen. In den derzeit durchgeführten Studien spielt der Novelty-Effekt eine große Rolle. Die Validität der Studienergebnisse ist deshalb eine andere als in bereits etablierten Fachgebieten.

Auf dem Gebiet der Aufmerksamkeitslenkung gibt eine reichliche Anzahl an Veröffentlichungen (Aitamurto u. a., 2018; Gugenheimer u. a., 2016b; Koskinen u. a., 2017; Y.-T. Lin u. a., 2017; Mateer, 2017; Nielsen u. a., 2016; Pillai u. a., 2017; Sassatelli u. a., 2018), welche in dieser Arbeit klassifiziert und einer einheitlichen Terminologie zugeordnet werden (Kapitel 2).

Wir hoffen, dass diese Arbeit einen Beitrag zur weiteren Entwicklung dieses Forschungsgebietes leisten kann.

2

How to Guide: Aufmerksamkeitslenkung in CVR

„Die Pioniere hatten keine Wegmarkierungen, die ihnen die Richtung zu der gesuchten Filmsprache gewiesen hätten. Viele der Konzepte, die entwickelt wurden, waren so kopflastig und abstrakt, dass kein Bezug zur Realität zu erkennen war. Doch diese Wegbereiter des Films waren ungeachtet all ihrer Fehler, Selbsttäuschungen und fehlgeleiteten Entdeckungen gewissenhafte Leute. Wenn die von ihnen aufgestellten Regeln überhaupt einen praktischen Wert besitzen, dann den, dass sie das Ergebnis systematischen Experimentierens sind, eine Ansammlung von Lösungen, die sich aus der täglichen Ausübung ihres Handwerks ergaben.“

Daniel Arijon (Arijon, 2000)

Filmische Elemente wie Töne, Lichter und Bewegungen lenken die Aufmerksamkeit des Betrachters im Film. In CVR schaut der Betrachter sphärische Filme mithilfe von HMDs oder anderen VR-Geräten an. Dabei ist die Blickrichtung frei wählbar, so dass wichtige Details möglicherweise außerhalb des Sichtfelds des Betrachters liegen. Für einige CVR-Experiences erweist sich dies als unproblematisch: Der Benutzer entdeckt eine vom Autor erstellte Storywelt, die keine für den Handlungszeitpunkt besonders ausgewiesenen Bereiche aufweist und damit keine zusätzlichen Guiding Methoden erfordern. In anderen Story-Konstrukten ist es wichtig, bestimmte Details nicht zu übersehen, und die Zuschauer sollten auf diskrete Weise geführt werden, damit sie entspannt die Applikation genießen können. In solchen Fällen können Guiding Methoden die User Experience verbessern.

Dieses Kapitel enthält bereits veröffentlichtes Material aus folgenden Publikationen:

- Rothe, S., Buschek, D., Hußmann H., 2019. Guidance in Cinematic Virtual Reality-Taxonomy, Research Status and Challenges. *Multimodal Technologies and Interaction* 3
- Rothe, S., Hußmann, H., 2018. Guiding the Viewer in Cinematic Virtual Reality by Diegetic Cues. in: *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*
- Rothe, S., Althammer, F., Khamis, M., 2018. GazeRecall: Using Gaze Direction to Increase Recall of Details in Cinematic Virtual Reality. in: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia - MUM'18*
- Rothe, S., Hußmann, H., Allary, M., 2017. Diegetic cues for guiding the viewer in cinematic virtual reality. in: *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17*

CVR-Anwendungen werden in unterschiedlichen Bereichen verwendet: Unterhaltung, Kunst, Bildung, Werbung oder sogar Anleitungen und Tutorials. Wie viel und welche Art der Führung des Betrachters benötigt wird, hängt in hohem Maße von der Zielvorstellung und dem Filminhalt ab. In einigen Fällen ist sie für die Kontinuität der Geschichte erforderlich, in anderen für Interaktionen (Abschnitt 4.2), Untertitel (Abschnitt 5.3), oder auch soziale Kommunikation (Abschnitt 5.1). Oft werden in der Literatur die beiden Begriffe *Aufmerksamkeitslenkung* (engl.: *Attention Guidance*) und *Blicklenkung* (engl.: *Gaze Guidance*) gleichartig verwendet (Booth u. a., 2013; McNamara u. a., 2012; Vig u. a., 2011). Allerdings muss gelenkte Aufmerksamkeit nicht immer visueller Natur sein und eine Blicklenkung muss nicht immer auch die Aufmerksamkeit dirigieren. In diesem Kapitel beschäftigen wir uns mit Methoden, deren Ziel es ist Blick und Aufmerksamkeit auf eine Rolle zu lenken und bezeichnen solche Methoden als *Guiding Methoden*.

In den letzten Jahren wurden mehrere Ansätze für Guiding Methoden in CVR veröffentlicht (Danieau u. a., 2017; Y.-C. Lin u. a., 2017; Sheikh u. a., 2016). Da CVR viele Gemeinsamkeiten mit anderen XR-Bereichen aufweist, haben wir uns auch mit Konzepten dieser Gebiete auseinandergesetzt. Ebenfalls sind Methoden für audiovisuelle Inhalte auf Flachbildschirmen und Mobilgeräten relevant. In all diesen Bereichen werden seit Jahren verschiedene Guiding Techniken untersucht und verwendet. Zum besseren Verständnis dieser Techniken erläutern wir in Abschnitt 2.1 einige Grundbegriffe der Aufmerksamkeitstheorie und Augenphysiologie und geben einen Überblick über die veröffentlichten Methoden.

Weitere Forschungsarbeiten werden in den nächsten Jahren erwartet, und dies erfordert einheitliche Begriffe für die Diskussion dieser Techniken. Zudem ist die Klassifizierung der Konzepte auch für das Finden neuer Ansätze hilfreich. Um zu diskutieren, welches Attribut einer Methode für den Erfolg oder Misserfolg relevant war, ist eine einheitliche übergreifende Terminologie erforderlich. Mit unserer Taxonomie in Abschnitt 2.2 leisten wir einen Beitrag zur Strukturierung der Methoden und damit zur weiteren Erforschung von Guiding in CVR.

Bei der Anwendung dieser Taxonomie in Abschnitt 2.3 unterscheiden wir zwischen verschiedenen Medien. Methoden, die beim herkömmlichen Filmmachen oder für Bilder verwendet werden, können in CVR angewendet werden, um den Betrachter im aktuellen Sichtfeld zu führen, wie in Abschnitt 2.3.1 beschrieben wird. VR und AR sowie CVR haben zusätzliche Anforderungen an Guiding Methoden, da sich relevante Objekte auch außerhalb des Sichtfeldes befinden. Diese Methoden werden in Abschnitt 2.3.2 beschrieben. Um mit unseren bisherigen Erfahrungen Entwicklern von CVR-Applikationen Hilfestellungen in der Verwendung der Guiding Methoden zu geben, wird in Abschnitt 2.3.3 aufgezeigt, wie die unterschiedlichen Attribute der Taxonomie für die bewusste Gestaltung genutzt werden können.

Am Ende dieses Kapitels diskutieren wir, wie die eingeführte Taxonomie den Designprozess von Guiding Methoden in CVR unterstützt. Die Taxonomie fördert das Verständnis der verschiedenen Attribute von Guiding Techniken, um neue Methoden zu finden und Filmemacher bei der Auswahl der richtigen Methoden für ihre Projekte zu unterstützen.

2.1 Hintergrund

Wissenschaftler aus verschiedenen Bereichen wie Psychologie, Biologie und Informatik arbeiten an Themen rund um Aufmerksamkeits- und Blicklenkung (Frintrop u. a., 2010). Grundkenntnisse in diesen Bereichen sind erforderlich, um Guiding Methoden zu verstehen. Aus der Filmtheorie übernehmen wir dafür die Begriffe diegetisch und nicht-diegetisch. *Diegetische* Cues sind Teil der Szene - zum Beispiel ein Musiker, der Musik spielt. *Nicht-diegetische* Cues kommen von außen - zum Beispiel Filmmusik oder ein Voice-Over.

2.1.1 Aufmerksamkeitstheorie / Attention Theory

Es gibt mehrere Faktoren, die beeinflussen wohin eine Person schaut. Auf der einen Seite gibt es *bottom-up* Faktoren, die die Szene charakterisieren. Dies sind Reize, die durch ihre Eigenschaften wie Farbe oder Form auffallen und normativ sind. Methoden sind *normativ*, wenn sie für alle Menschen gleich funktionieren, es sei denn eine Person leidet unter einer bestimmten Erkrankung wie beispielsweise Farbblindheit. Auf der anderen Seite gibt es *top-down* Faktoren wie Aufgaben oder Ziele. Die Wirksamkeit solcher Faktoren kann zwischen Individuen variieren. Je nach Ziel kann die Aufmerksamkeit *raumbasiert* (Position eines Objekts), *objektbasiert* oder *merkmalbasiert* (Merkmale eines Objekts) sein (Borji und Itti, 2013; Frintrop u. a., 2010).

Filmszenen können auf verschiedene Weisen erkundet werden: stimulusgesteuert (*bottom-up*) oder aufgabengesteuert (*top-down*). Die Betrachter können mithilfe von Lichtern, Farben und Tiefenschärfe durch Inszenierungs- und Kompositionstechniken geführt werden (*bottom-up*) aber auch durch das Verfolgen der Geschichte und Charaktere (*top-down*). Diese Prozesse sind dafür verantwortlich, dass Zuschauer Schnitte in einem Film oft nicht wahrnehmen. Die Aufgabe, der Geschichte zu folgen, verursacht die Edit-Blindheit, die Schnitte unsichtbar macht (Smith und Henderson, 2008). Dieser Effekt könnte auch für Guiding Methoden nützlich sein.

Cues können die Aufmerksamkeit auf ein Ziel lenken. Sie können verschiedene Eigenschaften und Positionen haben. Posner (Posner, 1980) zeigte, dass Betrachter ein Ziel schneller finden, wenn der Cue ein Merkmal des Ziels ist (z. B. ein farbiger Rand). Ein Cue, der nicht auf dem Ziel positioniert ist (z. B. ein Pfeil in Richtung des Ziels) benötigt mehr Zeit in der Verarbeitung. Posner führte die Begriffe *exogen* und *endogen* ein. *Exogene* Cues sind stimulusgesteuert und wirken automatisch, zum Beispiel ein Lichtblitz, der die Aufmerksamkeit auf sich zieht. Solche Cues verursachen eine reflexartige Orientierung, sind auf dem Ziel positioniert und können auch auditiv oder haptisch sein (Ward, 2008). Sie arbeiten als *bottom-up* Prozess. Da die Reaktion auf solche Cues reflexartig ist, wirken sie schnell. Wenn es jedoch keine interessante Zielinformation gibt, ist die Aufmerksamkeit nur kurzzeitig. *Endogene* Cues sind zielgerichtet und freiwillig (Ward, 2008). Oft basieren sie auf einem Zeichen, welches angibt, wo hingeschaut werden soll, und erfordern zunächst eine Interpretation. Selbst wenn das zielgesteuerte Guiding langsamer arbeitet, verbessert es die Verarbeitung des Ereignisses (Posner, 1980) und die Information kann für längere Zeiträume aufrechterhalten bleiben. Yarbus (Yarbus, 1967) zeigte, dass die Art der Augenbewegungen von der Aufgabe abhängt. In seinem Experiment sahen die Teilnehmer die gleiche Szene, nachdem ihnen verschiedene Fragen gestellt worden waren. Die Augenbewegungen unterschieden sich signifikant.

Eines der einflussreichsten Modelle der menschlichen visuellen Aufmerksamkeit, die *Merkmalsintegrationstheorie* (engl: Feature Integration Theory), wurde von Treisman und Gelade entwickelt (Treisman und Gelade, 1980). Sie erklärt die Rolle der visuellen Aufmerksamkeit für die Objekterkennung. Wird ein Reiz wahrgenommen, werden Eigenschaften frühzeitig, automatisch und parallel registriert. Objekte werden später in einem separaten Prozess identifiziert. Im ersten Schritt, der *präattentiven* Phase, sammeln Teile des Gehirns automatisch Informationen zu Merkmalen wie Farbe und Form. Im zweiten Schritt, der *attentiven* Verarbeitung, wird das gesamte Objekt durch Kombination der einzelnen Merkmale wahrgenommen.

Exogene und *präattentive* Prozesse sind unbewusst, subtil und werden durch Cues auf dem Ziel verursacht (direkte Cues). Im Gegensatz dazu sind *endogene* und *attentive* Prozesse bewusst und zielgerichtet. Dabei werden Merkmale analysiert und interpretiert. *Direkte* Cues können auffällige Merkmale eines Objekts sein. Healey et al. (Healey u. a., 1996) veröffentlichten eine Liste von zweidimensionalen auffälligen Merkmalen, ergänzt durch Literatur, die diese Merkmale verwenden. Einige für CVR relevante Beispiele sind: Farbe, Größe, Krümmung, Linienorientierung, Intensität, Flickern, Bewegungsrichtung und Beleuchtungsrichtung. Wolfe und Horwitz (Wolfe und Horowitz, 2017, 2004) beschrieben ebenfalls Zielattribute, die die Aufmerksamkeit effizient lenken können: Farbe, Orientierung, Größe, Tiefe, Bewegung und Leuchtdichte.

Die oben genannten Begriffe sind alle für die Charakterisierung des Aufmerksamkeitsprozesses relevant. Sie beschreiben verwandte, aber unterschiedliche Aspekte der Aufmerksamkeitslenkung und sind nicht orthogonal zueinander. Tabelle 2-1 gibt einen Überblick ohne Anspruch auf Vollständigkeit. Für unsere Taxonomie haben wir die Cue-Eigenschaft (letzte Zeile der Tabelle) ausgewählt, da diese unsere anderen extrahierten Dimensionen am besten ergänzt.

Prozess	bottom-up	top-down
Orientierung	exogen	endogen
Impuls	stimulusgesteuert	aufgabengesteuert
Automatismus	automatisch/reflexiv	freiwillig
Attentiveness	präattentiv	attentiv
Wahrnehmung	unbewusst	bewusst
Awareness	subtil	apparent
Cues	direkte Cues	indirekte Cues (Symbole)

Tabelle 2-1: Attribute von Guiding Methoden, die in der Literatur verwendet werden.

Das Wissen über die Forschungsergebnisse der Psychologie zur Erklärung der Aufmerksamkeitslenkung ist von grundlegender Bedeutung für das Verständnis der verschiedenen Guiding Methoden in der Human-Computer-Interaktion und beim Filmemachen. Je nach Anwendung oder Filmgenre können verschiedene Methoden geeignet sein.

2.1.2 Grundlagen der Physiologie der Augen

Für das Studium der Blicklenkung sind Kenntnisse der Augenphysiologie notwendig. Das menschliche Auge verfügt über zwei Arten von Sinneszellen: Stäbchen und Zapfen. Signale wie Farben oder Flicker werden in den Augenregionen unterschiedlich wahrgenommen. In der Peripherie befinden sich die Stäbchen, die für das Sehen in der Dunkelheit verantwortlich und für Licht und Bewegung sehr empfindlich sind. Das bedeutet, dass Cues an der Peripherie flackerndes Licht oder sich bewegende Elemente nutzen können. Die Zapfen in der Fovea (Bereich des schärfsten Sehens) werden benötigt, um Farben zu sehen. Sie reagieren weniger auf Licht. Auffällige Farben könnten verwendet werden, um die Aufmerksamkeit auf ein Objekt zu lenken, über das der Betrachter den Blick schweifen lässt.

Die unterschiedlichen Eigenschaften von Peripherie und Fovea sind der Grund für die unterschiedliche Wahrnehmung von Flickern in Abhängigkeit von der Blickrichtung. Die Critical Flicker Fusion Frequency (CFF) ist die Rate, mit der das Flickern fusioniert und als kontinuierlich wahrgenommen wird. Sie beträgt bei schlechten Lichtverhältnissen etwa 22 bis 25 Hz. Bei höherer Lichtintensität erhöht sich die CFF mit dem Logarithmus der Lichtintensität (Ferry-Porter-Gesetz) (Tyler und Hamer, 1993) je nach Größe auf bis zu 80 Hz (Granit-Harper-Gesetz) (Rovamo und Raninen, 1988). Je größer der flickernde Reiz ist, desto höher ist die CFF.

Dies ist der Grund für unterschiedliche CFFs in verschiedenen Regionen des Auges. Somit kann ein hochfrequenter Flicker in der Peripherie sichtbar sein, jedoch in der Fovea fusioniert. Zusätzlich erhöht sich die zeitliche Auflösungsschärfe (temporal resolution acuity: kleinstes Zeitintervall, das als Flicker wahrgenommen werden kann) für größere Flicker (Grimes, 1983). Diese Eigenschaft kann für eine subtile Blicklenkung verwendet werden. Die CFF ist jedoch eine sehr empfindliche Charakteristik und hängt nicht nur von den Regionen des Auges ab, sondern auch von den folgenden Parametern (Grimes, 1983; Rovamo und Raninen, 1988):

- Frequenz der Modulation
- Region auf der Netzhaut, die die Modulation wahrnimmt
- Amplitude / Tiefe der Modulation
- Leuchtdichte (Ferry-Porter-Gesetz) (Tyler und Hamer, 1993)
- Größe der Modulation (Granit-Harper-Gesetz) (Rovamo und Raninen, 1988)
- Farbe
- Kontrast
- Person (Alter, Müdigkeit)

Die CFF kann verwendet werden, um Stimuli-Cues zu erstellen und Guiding Methoden zu entwickeln. Es wurden einige Untersuchungen durchgeführt, bei denen die CFF zur Blicklenkung in Bildern verwendet wurde (Waldin u. a., 2017). Solche Methoden können jedoch nicht ohne Weiteres an Filme angepasst werden, da Veränderungen im Filmbild die Flickerreize unwirksam machen können und die Schwellwerte erhöht werden müssen, wodurch die Methode nicht mehr subtil ist (**Studie 7.1.5**) (Rothe u. a., 2018a). Auch wenn es schwierig ist, alle genannten Parameter auf einmal zu berücksichtigen und für jede Person gültige CFF-Schwellwerte zu finden, muss die Kenntnis über dieses Verhalten bei der Analyse und Entwicklung von Guiding Methoden berücksichtigt werden. Flicker können nicht nur als Teil einer Methode entworfen werden, sondern sie können auch in den Film einbezogen werden (z. B. flackernde Lichter). Möglicherweise helfen neue Technologien wie Künstliche Intelligenz dabei, dynamisch anpassbare Schwellwerte für Flicker-Methoden zu finden, so dass diese sowohl subtil als auch wirksam sind.

2.1.3 Aufmerksamkeitslenkung im Film

Das Konzept der Diegesis in der Filmtheorie wurde von Souriau entwickelt (Souriau, 1951) und anschließend auch in anderen Gebieten z. B. der Literaturtheorie verwendet. Diegetische Elemente gehören zur narrativen Welt. Das Konzept der Diegesis wird in der Filmtheorie oft für Musik und andere Klänge benutzt. Diegetische Musik in einem Film ist Teil der Geschichte. Sie kann nicht nur vom Zuschauer (wie Filmmusik), sondern auch von den Charakteren gehört werden. Beispiele sind: Musik aus einem Radio im Film oder Musik von Musikern, die Filmcharaktere sind.

Studien haben gezeigt, dass die Muster der Blickfixierungen bei Zuschauern von Filmen im Hollywood-Stil häufig konsistent sind (Dorr u. a., 2012; Subramanian u. a., 2014). In solchen Filmen wenden Filmemacher Erzähl- und Bearbeitungstechniken an, um den Betrachter zu wichtigen Aspekten einer Szene zu führen, häufig auf Kosten von peripheren Details (Subramanian u. a., 2014). Bilder in *framed* Filmen zeigen immer einen Ausschnitt, und im Rahmen, der Kadrierung (engl.: framing), ordnet der Filmemacher die Elemente des Films an. Untersuchungen mit Eyetrackern ergaben, dass Zuschauer die Peripherie des Filmbildes selten erkunden (Smith, 2013), wenn es keine Untertitel gibt. Menschen schauen meist auf den zentralen Bereich einer Kadrierung. In der Realität gibt es eine solche Kadrierung der Welt nicht, in CVR wird die Kadrierung vom Betrachter bestimmt. Daher braucht es in CVR neue Herangehensweisen.

2.1.4 Guiding Methoden

Die Guiding Methoden in Bildern, *framed* Filmen, VR, AR und CVR werden intensiv erforscht. In (Rothe u. a., 2019a) haben wir einen Überblick über veröffentlichte Methoden und die Umgebungen gegeben, in denen sie getestet wurden. Dabei wurden nicht nur Methoden berücksichtigt, die in CVR-Umgebungen evaluiert wurden, da es möglich ist, Techniken aus anderen Bereichen an CVR anzupassen. Einige dieser Methoden wurden von uns ausführlich evaluiert: diegetische Methoden in **Studie 7.1.1** und Subtle Gaze Direction (SGD) in den **Studien 7.1.4 - 7.1.6**. Es hängt von der Art des Films ab, ob Guiding Methoden erforderlich sind und wie stark oder subtil das Guiding sein sollte. Offensichtlichere Methoden sind effektiver, können jedoch das VR-Erlebnis beeinträchtigen (Danieau u. a., 2017). Da Guiding auch die Recall-Rate von Zielobjekten erhöhen kann (Bailey u. a., 2012; Rothe u. a., 2018a), ist auch der Zweck des Films relevant, um die geeignetste Technik zu finden. Eine Guiding Methode kann in einer wissensvermittelnden CVR-Anwendung zielführend sein, in einem meditativen Film jedoch störend.

Einige der wichtigsten Guiding Methoden verschiedener Forschungsbereiche werden im Folgenden näher untersucht:

2.1.4.1 Diegetische Methoden

Einige Forschungen der letzten Jahre konzentrierten sich auf diegetische Guiding Methoden in CVR (Nielsen u. a., 2016; Pausch u. a., 1996; Rothe und Hußmann, 2018; Sheikh u. a., 2016). *Diegetische Cues* sind Teil des Films, zum Beispiel sich bewegende Charaktere, Lichter oder Geräusche. Nielsen et al. (Nielsen u. a., 2016) verglichen eine diegetische Technik (Glühwürmchen) mit einer nicht-diegetischen (forcierte Rotation). Mithilfe eines Fragebogens stellten sie fest, dass der diegetische Cue (Glühwürmchen) hilfreicher war. Darüber hinaus zeigen die Ergebnisse, dass der nicht-diegetische Cue die Presence verringern kann. In den Experimenten von Syrett et al. (Syrett u. a., 2016) haben sich Zuschauer zum Teil durch die freie Wahl der Blickrichtung abgelenkt gefühlt und manchmal wichtige Teile der Handlung übersehen.

Nur wenige Forschungen haben sich damit befasst, wie man die Aufmerksamkeit des Betrachters in CVR durch diegetische Cues lenken kann. 1996 untersuchten Pausch et al. (Pausch u. a., 1996), wie die Aufmerksamkeit des Betrachters auf ein gewünschtes Detail gelenkt werden kann und verwendeten dafür die Charaktere als diegetische Cues. In ähnlicher Weise verknüpften Sheikh et al. (Sheikh u. a., 2016) mehrere diegetische Cues (Bewegungs-, Gestik- und Audio-Cues) mit der Hauptfigur einer Szene. Im Experiment erwiesen sich die Cues mit einer zusätzlichen Audiokomponente hilfreicher als reine visuelle Cues, auch bei nicht-räumlichem Sound. Die Hauptfigur eines Films zieht jedoch generell mehr Aufmerksamkeit auf sich, auch ohne zusätzliche Cues. In **Studie 7.1.1** wurde untersucht, ob es möglich ist, die Aufmerksamkeit anhand von Cues zu steuern, die mit neutralen Objekten verbunden sind, Objekten, die der Betrachter zuvor noch nicht gesehen hat.

Auf der anderen Seite gibt es nicht-diegetische Cues, die nicht Teil der Geschichte sind, wie z. B. hinweisende Pfeile oder forcierte Rotationsmethoden (C. Brown u. a., 2017; Silva u. a., 2003).

2.1.4.2 Saliency-Modulationstechniken (SMT)

Veas et al. (Veas u. a., 2011) beschrieben Saliency-Modulationstechniken (SMT), um die Aufmerksamkeit des Betrachters auf ein Zielobjekt zu lenken. Für die Analyse der Originalbilder wurden Saliency Maps verwendet und das Material in Abhängigkeit von den Ergebnissen moduliert. Eine Saliency-Map zeigt die Saliency-Werte (Auffälligkeiten) für jede Region im Bild an (Itti u. a., 1998; Niebur, 2007). Auf diese Weise konnten minimale Modulationen vorgenommen werden. Die vorgestellte Methode arbeitet mit Videos in Echtzeit. Sie kann jedoch nur verwendet werden, wenn sich die RoI im FoV des Betrachters befindet. Darüber hinaus ist die Methode in Umgebungen mit sich bewegenden und blinkenden Objekten weniger effektiv. Veas et al. (Veas u. a., 2011) untersuchten SMT hinsichtlich Modulations-Awareness, Aufmerksamkeit und Gedächtnis. Sie zeigten, dass SMT die Aufmerksamkeit auf ausgewählte Ziele lenken kann, ohne dass der Betrachter die Modulation bemerkt. Darüber hinaus kann SMT die Recall-Rate erhöhen.

Da Saliency-Modulationstechniken den Blick lenken können (Barth u. a., 2006; De Abreu u. a., 2017; Hagiwara u. a., 2011; Sitzmann u. a., 2018; Veas u. a., 2011; Vig u. a., 2011), könnten die Techniken auf CVR übertragbar sein. Die Saliency-Modulation kann jedoch nur dann wirksam sein, wenn sich der modulierte Bereich im Sichtfeld des Betrachters befindet.

2.1.4.3 Blurring / Unschärfe

Smith et al. (Smith und Tadmor, 2013) untersuchten Unschärfen, um die Aufmerksamkeit des Betrachters in Bildern zu lenken. Sie zeigten, dass der Betrachter zu Regionen mit geringer oder keiner Unschärfe tendiert, wenn der Rest des Bildes unschärfer ist. Dieser Ansatz ist Techniken der traditionellen Filmproduktion sehr ähnlich. Hata et al. (Hata u. a., 2016) haben diese Guiding Methode erweitert und unbemerkte Unschärfeneffekte in Bildern genutzt. Es wurden Schwellwerte gefunden, unter denen der Betrachter die Unschärfe nicht bemerkt, aber dennoch durch sie gelenkt werden kann.

2.1.4.4 Stylistic Rendering / Stilistisches Rendern

Die Lenkung des Betrachters in einem Film kann auch durch unterschiedliche Stilelemente erfolgen: Schärfentiefe, Farben und Helligkeit. Cole et al. (Cole u. a., 2006) untersuchten Guiding in 3D-Modellen. Sie verwendeten lokale Variationen von Effekten (Farbsättigung und Kontrast) und Eigenschaften (Textur und Dichte), um den Blick des Betrachters auf den hervorgehobenen Bereich zu lenken. Zusätzlich setzten sie eine dynamische Technik ein: das stilisierte Ändern des Fokus (stylized focus pull). Schärfeverlagerung (Focus Pull) ist eine kreative Kameratechnik im Film, bei der sich der Fokus während der Aufnahme ändert und die Aufmerksamkeit von einem Bereich zum anderen wechselt (Wulff, 2012a). Bei der digitalen Bearbeitung kann das Ziehen des Fokus durch einen animierten Filtereffekt auch in der Postproduktion hinzugefügt werden.

2.1.4.5 Subtle Gaze Direction (SGD) / Subtile Blicklenkung

SGD mit Eyetracking

Ziel einer subtilen Blicklenkungen ist es, für den Betrachter unbemerkt zubleiben. Das Konzept der subtilen Gaze Direction (SGD) wurde von Bailey eingeführt (Bailey u. a., 2009). Der Kern dieses Konzepts besteht darin, den Zielbereich zu modulieren, wenn er sich in der Peripherie befindet, um den Betrachter dazu zu bewegen, dorthin zu schauen, und die Modulation zu stoppen, wenn der Betrachter in diese Richtung blickt. Auf diese Weise kann der Blick des Betrachters gelenkt werden, ohne dass die Modulation wahrgenommen wird. In der Forschung von Bailey et al. (Bailey u. a., 2009), wurden zwei Optionen untersucht: Luminanzmodulation und Rot-Blau-Modulation, beide mit einer Frequenz von 10 Hz in einem kreisförmigen Bereich von etwa 1 cm Durchmesser. Ein Eyetracker wurde verwendet, um den Blick des Betrachters zu verfolgen. Die Modulation wurde gestoppt, wenn der Betrachter den Blick in Richtung des Zielbereichs änderte. Es konnte gezeigt werden, dass diese Technik den Blick des Benutzers in Standbildern effektiv lenken kann, ohne dass der Benutzer die Modulation bemerkt.

In den Experimenten von McNamara et al. (McNamara u. a., 2008) erwies sich die Modulation der Luminanz effektiver als die Rot-Blau-Modulation. Es wurde gezeigt, dass SGD die Ergebnisse von Suchaufgaben in Bildern verbessert, ohne dass die Teilnehmer die Modulation bemerken. Diese Methode wurde auch für Guiding in narrativer Kunst mit statischen Bildern untersucht (McNamara u. a., 2012). Grogorick et al. erweiterten die Methode auf virtuelle Umgebungen (Grogorick u. a., 2017). Eine Luminanzmodulation wurde eingesetzt und der Stimulus dynamisch positioniert, so dass die Methode auch für Ziele verwendet werden konnte, die sich zu Beginn der Stimulation nicht im FoV befinden. Die Experimente haben gezeigt, dass die Ergebnisse von Suchaufgaben verbessert werden können.

SGD mit hochfrequenten Flickern

Es gibt einige Untersuchungen dazu, wie SGD ohne einen Eyetracker verwendet werden kann. Waldin et al. (Waldin u. a., 2017) nutzten die Tatsache, dass das periphere Sehen empfindlicher für hochfrequentes Flickern ist als das foveale Sehen (Abschnitt 2.1.2). Daher ist die kritische Fusionsfrequenz (CFF) in diesen Bereichen unterschiedlich. Ein in der Peripherie flackerndes Signal wird im Bereich der Fovea nicht mehr

als flackernd wahrgenommen. Der Flicker fusioniert zu einem stabilen Bild. Auf diese Weise ist kein Eyetracking zum Stoppen der Modulation erforderlich. Die Experimente verwendeten Flicker von 60 Hz und 72 Hz, so dass ein hochfrequenter Monitor mit 120 Hz und 142 Hz erforderlich war. Im ersten Experiment wurden Bilder mit Kreisflächen verwendet, im zweiten Experiment ein hochkomplexes Bild. Die Methode funktionierte in beiden Fällen effektiv. Es scheint notwendig zu sein, eine persönliche Kalibrierungsroutine auszuführen, um die Größe und Luminanz der Flickermodulation zu ermitteln. Derzeit kann diese Methode nicht an VR und CVR angepasst werden, da die Frequenz der HMD-Anzeigen (90 Hz) nicht hoch genug ist. Außerdem ist der Flicker in Umgebungen mit dynamischen Änderungen weniger effektiv. Wie in Abschnitt 2.1.2 erwähnt, hängt die CFF nicht nur von der Augenregion ab, auch die Schwellwerte sind schwer zu finden. (**Studie 7.1.5**)

2.1.4.6 Off-screen Indikatoren

Da der Betrachter in CVR nur einen Ausschnitt des Filmbildes über HMD beobachtet, sind die oben genannten Methoden nicht immer anwendbar. Je nach Blickrichtung können Cues nicht gesehen werden, wenn sie nicht im FoV des Betrachters sind. Daher sind Methoden erforderlich, um auf Rols außerhalb des Screens zu verweisen. Wir nennen diese *off-screen* Indikatoren.

Eine Möglichkeit zur Visualisierung von Objekten außerhalb des Screens auf Flachbildschirmen ist die Halo-Technik (Baudisch und Rosenholtz, 1993). Dabei werden Objekte außerhalb des Screens von Kreisen umgeben, deren Größe ausreicht, um am Rand des Displays sichtbar zu sein. Aus der Position und Krümmung des Kreises kann der Benutzer die Position des Objekts ableiten.

Die Halo-Methode ist nicht direkt auf CVR übertragbar, da der CVR-Screen eine Sphäre ist. Damit am Rand des HMD-Displays noch ein Kreis sichtbar ist, darf dessen Mitte nicht weiter als 90° entfernt sein.

EdgeRadar (Gustafson und Irani, 2007) und Wedge (Gustafson u. a., 2008) sind Modifikationen dieser Technik, um Überladung und Überlappung zu vermeiden. Diese Techniken wurden an mobile AR (Perea u. a., 2017) und HMDs (Gruenefeld u. a., 2018a) angepasst. Gruenefeld et al. (Gruenefeld u. a., 2017a) verglichen mehrere Visualisierungstechniken für Objekte außerhalb des Screens in AR (Pfeil, Halo und Wedge). In ihren Experimenten schnitten die Halo- und Wedge-Technik am besten ab. Die implementierten Methoden waren jedoch auf einen 90°-Bereich vor dem Benutzer beschränkt und bedürfen einer weiteren Anpassung an CVR. EyeSee360 (Gruenefeld u. a., 2017b) ist eine Visualisierungstechnik für Objekte in AR, die für CVR angepasst werden kann.

2.1.4.7 Forcierte Rotation

Forcierte Rotation des Nutzers (SwiVRChair)

Gugenheimer et al. (Gugenheimer u. a., 2016b) entwickelten einen Stuhl, der den Betrachter automatisch dreht, um vordefinierte Rols zu zeigen. In ihrem Experiment war die Simulator-Sickness sehr gering. Dies wird durch die Tatsache verursacht, dass der Betrachter gedreht wurde und die Rotation in der virtuellen Welt mit der Rotation in der realen Welt übereinstimmt. Zusätzlich benötigten die Teilnehmer weniger Kopfbewegungen und konnten so das VR-Erlebnis entspannter genießen.

Forcierte Rotation der VR-Welt

Eine andere Möglichkeit der forcierten Rotation besteht darin, die Szene so zu drehen, dass sich Rol im Sichtfeld des Betrachters befindet. Nielsen et al. (Nielsen u. a., 2016) verglichen forcierte Rotation mit diegetischem Guiding. In ihren Experimenten war die diegetische Methode hilfreicher und resultierte in einer höheren Presence. In den Studien von Lin et al. (Y.-C. Lin u. a., 2017) wurde die forcierte Rotation (dort Autopilot genannt) mit einem Pfeil verglichen, der in Richtung der Rol zeigt (visuelles Guiding genannt). Beide Methoden halfen dem Betrachter, die Rol zu sehen. Aus den Ergebnissen ging hervor, dass die bevorzugte Methode von den Interessen des Zuschauers und dem Inhalt des Films abhängt. Es wurde jedoch keine allgemein höhere Simulator-Sickness für die forcierte Rotation beobachtet. Eine Ursache von Simulator-Sickness ist die Diskrepanz zwischen Bewegungen in der realen und der virtuellen Welt (Davis u. a., 2014; Kolasinski, 1995), weshalb das Rotieren der VR-Welt vor dem Benutzer häufig zu

Unwohlsein führt. Es gibt keine einheitliche Meinung darüber, ob das Rotieren einer Szene zur Simulator-Sickness führt oder nicht (Y.-C. Lin u. a., 2017).

Forcierte Rotation durch Schnittmethoden

Beim herkömmlichen Filmmachen kann der Schnitt genutzt werden, um dem Betrachter zu wichtigen Details zu führen: Nach dem Schnitt wird die Rol angezeigt. Das Gleiche kann in CVR angewendet werden: Unabhängig von der Blickrichtung vor dem Schnitt wird dem Betrachter die Rol nach dem Schnitt angezeigt (Pavel u. a., 2017; Sassatelli u. a., 2018) (**Studien 7.3.1 und 7.3.2**). Es muss jedoch untersucht werden, ob dies zu einer Desorientierung führen kann, wenn sich beide Szenen am selben Ort befinden und sich die Blickrichtung mit dem Schnitt ändert - ähnlich wie beim Achsensprung (Wulff, 2012b) in framed Filmen (**Studie 7.3.6**).

2.1.4.8 Auditive Cues

Die meisten der bisher beschriebenen Methoden beruhen auf visuellen Merkmalen. Diese haben allerdings den Nachteil, dass sie nur wirksam sind, wenn sie sich im FoV des Betrachters befinden. Auditive Cues können jedoch bei einer beliebigen Blickrichtung wahrgenommen werden. Van der Burg et al. (Van der Burg u. a., 2008) haben gezeigt, dass Audio-Cues (pip), die mit einem auffallendem visuellen Cue (pop) synchronisiert sind, die Suchzeit verkürzen, auch wenn der Audio-Cue keine Standortinformationen enthält. Dies entspricht auch einem der Ergebnisse aus **Studie 7.1.1**. Emil R. Høeg et al. (Hoeg u. a., 2017) erweiterten dieses Experiment auf VR mit Audio-Cues aus der Richtung des visuellen Cues. Sie zeigten, dass binaurale Cues zu kürzeren Suchzeiten führen, obwohl der visuelle Cue zum Zeitpunkt der Präsentation des Audio-Cues nicht unbedingt sichtbar war. In den Experimenten erhielten die Teilnehmer eine Suchaufgabe in einer abstrakten VR-Umgebung. Trotzdem sind diese Ergebnisse auch für CVR relevant, wo beim Filmschauen keine konkrete Aufgabe gegeben ist und kinematografische Szenen verwendet werden.

2.1.4.9 Haptische Cues

Um auch haptische Cues für das Guiding zu nutzen, wurden verschiedene Techniken entwickelt. *HapticHead*, eine Kappe, in die mehrere vibrotaktile Aktuatoren integriert sind, ermöglicht ein haptisch verursachtes Guiding (Kaul und Rohs, 2017). Chang et al. (Chang u. a., 2018) führten *FacePush* ein, ein System für haptische Signale in HMDs. Es erzeugt Kräfte im Gesicht des Betrachters und wurde für zwei VR- und eine CVR-Anwendung getestet. Im Gegensatz zu *HapticHead* oder einem vibrotaktilem Stirnband (Rantala u. a., 2017), zeigt *FacePush* die empfohlene Drehrichtung (links/rechts) und nicht die absolute Richtung der Rol an. Um haptische Signale als diegetische Signale in eine Geschichte zu integrieren, sind haptische Reize auf andere Körperteile jenseits des Kopfes erforderlich. Solche haptischen Reize können durch Drohnen ausgelöst werden (Hoppe u. a., 2018; Knierim u. a., 2018).

In der beschriebenen Literatur findet man verschiedene Eigenschaften von Guiding Methoden, z. B. subtil, off-screen, forciert, diegetisch, haptisch und einige andere. Nicht alle von ihnen sind miteinander vergleichbar, da sie verschiedene Aspekte der Methode hervorheben. Es ist wichtig, diese Attribute zu klassifizieren, um die relevantesten und qualifiziertesten Merkmale für Guiding Methoden in CVR zu finden. Wir werden dies im nächsten Abschnitt tun.

2.2 Taxonomie

Um geeignete Guiding Techniken für CVR zu finden, wurden Methoden für verschiedene Medien untersucht: Bilder, Filme, VR und AR (Abschnitt 2.1.4). Inspiriert von der Vielzahl von Artikeln und verschiedenen Taxonomien in VR (Nielsen u. a., 2016; Nilsson u. a., 2016; Suma u. a., 2012), haben wir diese Methoden analysiert und ihre Eigenschaften klassifiziert.

Mit unserer Klassifikation fanden wir sieben orthogonale Dimensionen. Nielsen et al. (Nielsen u. a., 2016) beschrieben drei Dimensionen der Aufmerksamkeitslenkung. Eine unserer Dimensionen (Diegesis) ist konsistent zu Nielsen et al., zwei andere entsprechen ihrer Taxonomie (Cue und Freedom).

Unsere Taxonomie beschreibt die wichtigsten Attribute, die wir in der Literatur fanden und die für unsere eigene Arbeit relevant sind, ohne Anspruch auf Vollständigkeit. Es ist denkbar, dass in Zukunft je nach Forschungsschwerpunkt neue Komponenten hinzukommen. Tabelle 2-2 zeigt unsere Taxonomie wichtiger Dimensionen und deren mögliche Werte. Diese werden in den folgenden Unterabschnitten erläutert.

Dimension	Attribute		
Diegesis	diegetisch	nicht-diegetisch	
Sinn (sense)	visuell	auditiv	haptisch
Ziel (target)	on-screen	off-screen	
Referenz	world-referenziert	screen-referenziert	
Cue	direkt	indirekt	
Awareness	subtil	apparent	
Freedom	forciert durch das System	forciert durch Reflexe	freiwillig

Tabelle 2-2: Die Tabelle zeigt die verschiedenen Dimensionen der Guiding Methoden und mögliche Werte.

2.2.1 Diegetisch und nicht-diegetisch

Forschungsergebnisse zeigen, dass *diegetische* Methoden in CVR gut funktionieren (Nielsen u. a., 2016; Rothe und Hußmann, 2018; Sheikh u. a., 2016). Für *visuelle diegetische* Techniken muss der Cue, z. B. Bewegungen, Licht, Farbe, im FoV liegen. In den meisten Fällen ist der Ort des Cues (z. B. die Farbe des Ziels) mit dem des Ziels identisch. Eine Ausnahme ist: Ein Protagonist schaut oder zeigt in eine bestimmte Richtung.

Man kann sich jedoch Szenen vorstellen, in denen keine geeigneten Cues zur Verfügung stehen. Wenn es dennoch erforderlich ist, die Aufmerksamkeit auf ein Detail zu lenken, können *nicht-diegetische* Methoden angewendet werden. Je nach Anwendungsfall muss die Methode so gestaltet sein, dass sie gut wahrgenommen werden kann oder möglichst unauffällig ist. Einige Vor- und Nachteile der diegetischen und nicht-diegetischen Methoden sind in Tabelle 2-3 aufgeführt.

	diegetisch	nicht-diegetisch
+	hohe Presence und Enjoyment	leicht anwendbar und erkennbar
-	abhängig von Story, nicht immer anwendbar	kann das VR-Erlebnis beeinträchtigen
->	geeignet für explorative Story-Strukturen	geeignet für wichtige Informationen

Tabelle 2-3: Vor- und Nachteile (+ und -) von diegetischen und nicht-diegetischen Methoden. Die dritte Zeile gibt an, für welche Anforderungen die Methoden geeignet sind.

2.2.2 Visuell, auditiv und haptisch

Es ist naheliegend, *visuelle* Guiding Methoden in CVR zu diskutieren. Bewegungen, Licht und Charaktere sind dafür bekannt, die Aufmerksamkeit in framed Filmen auf sich zu ziehen (Bordwell und Thompson, 2013). Diese Cues wirken jedoch nur, wenn sie sich im FoV befinden. Um den Benutzer zu motivieren, die

Blickrichtung zu ändern, sind *auditive* Cues aus der Richtung der RoI eine effektive Methode, da sie auch für Regionen außerhalb des Sichtfelds genutzt werden können (**Studien 7.1.1 und 7.1.2**). Auch wenn die Quelle eines Tons nicht sichtbar ist, ist es möglich, ihn zu hören - einschließlich seiner Richtung. Das entspricht der Erfahrung aus dem realen Leben: Ein Geräusch kann jemanden dazu bringen, die Blickrichtung zu ändern. Gleiches gilt für CVR (Rothe und Hußmann, 2018). Auch *haptische* Cues können dieses Verhalten verursachen, und es ist naheliegend, diese als Guiding Methode zu diskutieren. Einige Vor- und Nachteile von visuellen, auditiven und haptischen Methoden sind in Tabelle 2-4 aufgeführt.

	visuell	auditiv	haptisch
+	kann leicht integriert werden	auch für off-screen Rols	neue Erfahrung
-	nicht immer sichtbar, abhängig von Blickrichtung	schwer unterscheidbar zwischen diegetisch und nicht-diegetisch	schwer umzusetzen, zusätzliche Geräte notwendig
->	geeignet für explorative Story-Strukturen	geeignet für Änderung der Blickrichtung	geeignet für öffentliche Installationen

Tabelle 2-4: Vor- und Nachteile (+ und -) von visuellen, auditiven und haptischen Methoden. Die dritte Zeile gibt an, für welche Anforderungen die Methoden geeignet sind.

2.2.3 On- und Off-Screen

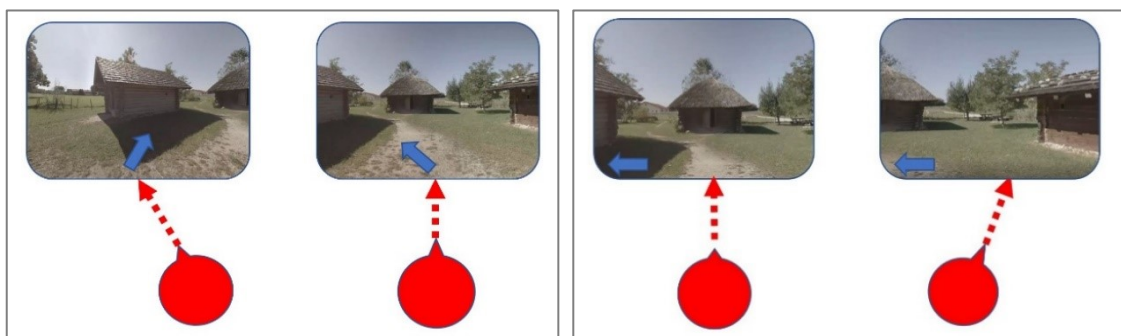


Abbildung 2-1: Beispiele für *on-* und *off-screen* Guiding. Links: *on-screen*. Der Pfeil verweist auf ein Detail auf der Hüttenwand, welches auf dem Display zu sehen ist. Dreht der Nutzer seinen Kopf und das Detail ist noch im Sichtfeld, ändert sich der Pfeil und zeigt weiterhin auf das Detail. Rechts: *off-screen*. Die RoI befindet sich außerhalb des Sichtfeldes und der Pfeil zeigt die Richtung an, in die der Kopf bewegt werden sollte.

Je nach Blickrichtung kann sich eine RoI im FoV des Betrachters oder außerhalb davon befinden. Um die Aufmerksamkeit auf ein Objekt im momentanen Sichtfeld zu lenken, können Methoden verwendet werden, die bereits für Bilder oder framed Filme untersucht wurden. Wir nennen dies *on-screen* Guiding (Abbildung 2-1, links). Visuelle Methoden, wie die Saliency-Modulation, können nur funktionieren, wenn sich die Region im FoV befindet. In CVR kann es jedoch vorkommen, dass der Betrachter zuerst die Blickrichtung ändern muss, um die RoI im FoV zu haben. In diesem Fall sind *off-screen* Methoden erforderlich (Abbildung 2-1, rechts).

Welche der beiden Methoden angewendet wird, hängt nicht vom Autor, sondern von der Blickrichtung ab. Der Autor muss entscheiden, ob beide benötigt werden.

2.2.4 World- und Screen-referenziert

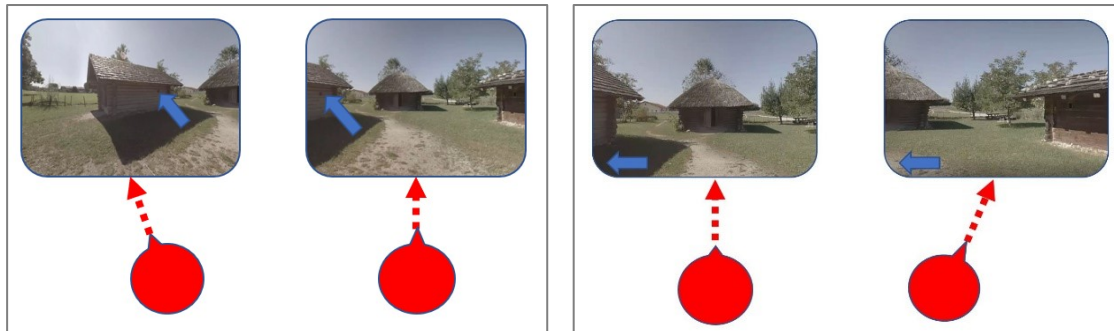


Abbildung 2-2: Beispiele für *world-* und *screen-referenziertes* Guiding. Links: world-referenziert. Der Pfeil verweist auf ein Detail auf der Hüttenwand, welches auf dem Display zu sehen ist. Dreht der Nutzer seinen Kopf bleibt der Pfeil in Position und Richtung, verankert mit der Filmwelt. Rechts: screen-referenziert. Der Pfeil bleibt immer am unteren Bildschirmrand, auch wenn der Kopf bewegt wird.

Cues in VR können in *screen-referenziert* und *world-referenziert* unterteilt werden (Shah und Miyake, 2005; Yeh u. a., 1999). *World-referenzierte* Objekte sind mit der virtuellen Welt verbunden, in unserem Fall mit dem Film (Abbildung 2-2, links). Sie bleiben an ihrem Platz in der Filmwelt fixiert, auch wenn der Betrachter den Kopf bewegt. *Screen-referenzierte* Elemente sind mit dem Display verbunden und bewegen sich mit, falls der Betrachter den Kopf dreht (Abbildung 2-2, rechts). Der Begriff *screen-referenziert* entspricht dem in AR verwendeten Begriff „*in-view*“ und *world-referenziert* dem Term „*in-situ*“ (Renner und Pfeiffer, 2017a).

Auch wenn sich diegetische Cues auf die Welt beziehen, müssen world-referenzierte Cues nicht notwendig *diegetisch* sein. Ein Cue, der in der Postproduktion hinzugefügt wird (z. B. ein Pfeil), um den Betrachter zu führen, der aber nicht Teil des Films ist und von den Filmcharakteren nicht wahrgenommen werden kann, ist nicht-diegetisch. Screen-referenzierte Cues sind immer nicht-diegetisch, da sie nicht Teil der (Film)-Welt sind. Einige Vor- und Nachteile von world-referenzierten und screen-referenzierten Methoden sind in Tabelle 2-5 aufgeführt.

	world-referenziert	screen-referenziert
+	integriert in VR-Welt, hohe Presence	immer sichtbar
-	nicht immer sichtbar	beeinträchtigt das VR-Erlebnis
->	geeignet für Anwendungen, in denen Presence wichtig ist	geeignet für off-screen Guiding und wichtige Informationen

Tabelle 2-5: Vor- und Nachteile (+ und -) von world-referenzierten und screen-referenzierten Methoden. Die dritte Zeile gibt an, für welche Anforderungen die Methoden geeignet sind.

2.2.5 Direkte und indirekte Cues



Abbildung 2-3: Links: ein direkter Cue auf der Region des Hauses. Rechts: ein Pfeil als indirekter Cue.

Es gibt zwei Haupttypen von Cues: *direkte* und *indirekte* Cues (Wright, 1998). *Direkte* Cues sind Teil des Zieles, z. B. Umrisse, Farben oder Lichter (Abbildung 2-3, links). *Indirekte* Cues basieren auf symbolischen Informationen und müssen zuerst interpretiert werden, z. B. ein Pfeil (Abbildung 2-3, rechts). Die Cues müssen nicht *visuell* sein. Zum Beispiel kann ein plötzliches Geräusch als *direkter auditiver* Cue und eine Stimme, die sagt, was gesehen werden kann, als *indirekter* Cue wirken.

Direkte Cues arbeiten meistens *stimulusgesteuert*, basieren auf den Eigenschaften der Szene, z. B. einem abrupten Licht oder einer plötzlichen Bewegung, und sind *bottom-up* Prozesse. Rols müssen sich dazu ausreichend von der Umgebung unterscheiden. *Direkte* Cues wirken schnell, vorübergehend und spontan (Borji und Itti, 2013; Carrasco, 2011).

Indirekte Cues erfordern eine bewusste Anstrengung, z. B. die Interpretation eines Zeichens. Sie sind *top-down* Prozesse und nutzen kognitive Eigenschaften wie Wissen, Erwartung und Aufgabe. *Indirekte* Methoden sind langsam, nachhaltig und freiwillig (Borji und Itti, 2013; Carrasco, 2011; Itti und Koch, 2001). Einige Vor- und Nachteile von direkten und indirekten Cues sind in Tabelle 2-6 aufgeführt.

	direkte Cues	indirekte Cues
+	schnell	nachhaltig
-	nicht immer sichtbar	müssen interpretiert werden
->	geeignet für on-screen Guiding	geeignet zum Merken von Rols

Tabelle 2-6: Vor- und Nachteile (+ und -) von world-referenzierten und screen-referenzierten Methoden. Die dritte Zeile gibt an, für welche Anforderungen die Methoden geeignet sind.

2.2.6 Subtil und apparent

Ist dem Benutzer eine Methode nicht bewusst, wird die Methode als *subtil* bezeichnet. Im Gegensatz dazu werden *apparente* Techniken vom Benutzer bemerkt (Suma u. a., 2012). Es gibt mehrere subtile Guiding Methoden, die auf der Physiologie des Auges basieren. Der Begriff „Subtile Blicklenkung“ (Subtle Gaze Direction, SGD) ist für diese Methoden bereits etabliert. Er wird in der Literatur jedoch nicht einheitlich verwendet. Der Begriff *unterschwellig* ist ebenso für Reize gebräuchlich, die unterhalb der Schwelle der bewussten Wahrnehmung liegen (Hata u. a., 2016). Wie bereits erwähnt, hängen solche Schwellwerte (z. B. CFF) von mehreren Faktoren ab und variieren zwischen Menschen. Somit kann ein Cue für eine Person unterschwellig sein, für eine andere Person jedoch offensichtlich.

Subtilität kann auch anders erreicht werden. Beispiele sind diegetische Methoden, bei denen Elemente des Films den Blick lenken. Der Benutzer bemerkt den Cue, ist sich jedoch normalerweise der Guiding Eigenschaft nicht bewusst. Auch wenn subtile Techniken als Kontinuum definiert werden können,

stimmen wir mit Suma et al. (Suma u. a., 2012) in der Wahl einer dichotomischen Kategorisierung überein (subtil vs. apparent), wobei der Begriff „subtil“ ein Oberbegriff ist und „unterschwellig“ enthält.

Einige Vor- und Nachteile subtiler und apparter Methoden sind in Tabelle 2-7 aufgeführt.

	subtil	apparent
+	keine Beeinträchtigung der Experience	leicht erkennbar kann Recall-Rate verbessern
-	nicht immer effektiv	kann störend sein
->	geeignet für explorative Story-Strukturen	geeignet für Lerntasks

Tabelle 2-7: Vor- und Nachteile (+ und -) subtiler und apparter Methoden. Die dritte Zeile gibt an, für welche Anforderungen die Methoden geeignet sind.

2.2.7 System-forciert, reflex-forciert und freiwillig

Die meisten der bisher diskutierten Methoden sind freiwillig: Die Zuschauer können frei entscheiden, ob sie den Cues folgen oder die Szene selbstständig erkunden. Es können jedoch auch forcierte Methoden angewendet werden (Gugenheimer u. a., 2016b; Y.-C. Lin u. a., 2017; Nielsen u. a., 2016; Stratmann u. a., 2018). Dabei gibt es verschiedene Möglichkeiten. Einerseits kann der Betrachter rotiert werden (SwiVRChair (Gugenheimer u. a., 2016b)), was den Vorteil hat, dass der Betrachter die Drehbewegung spürt. Andererseits kann die VR-Welt/der Film gedreht werden (Y.-C. Lin u. a., 2017). Diese Methoden zwingen den Benutzer, auf technische Weise die Blickrichtung zu ändern. Forcierung kann auch mit Methoden erreicht werden, die auf den in Abschnitt 2 beschriebenen physiologischen Modellen basieren. Reize können den Betrachter dazu veranlassen, die Blickrichtung reflexartig und schnell zu ändern. Einige Vor- und Nachteile von forcierten und freiwilligen Methoden sind in Tabelle 2-8 aufgeführt.

	system-forciert	reflex-forciert	freiwillig
+	Rol wird immer gezeigt	schnell, kann in die Story integriert werden	freie Wahl der Blickrichtung
-	kann das VR-Erlebnis beeinträchtigen	nicht immer anwendbar	Rol kann verpasst werden
->	geeignet für wichtige und schnelle Richtungswechsel	geeignet für visuelle Erlebnisse und schnelle Reaktionen	geeignet für explorative Story-Strukturen

Tabelle 2-8: Vor- und Nachteile (+ und -) forcierter und freiwilliger Methoden. Die dritte Zeile gibt an, für welche Anforderungen die Methoden geeignet sind.

2.3 Verwendung der Taxonomie

Im vorherigen Abschnitt wurden die von uns identifizierten Dimensionen für Guiding Methoden in CVR und ihre möglichen Werte beschrieben. Tabelle 2-9 stellt Methoden aus der Literatur dar und listet ihre Attribute in der eingeführten Taxonomie auf. Die Tabelle zeigt, dass bisher wenig subtile Methoden untersucht wurden. Wir konnten nur eine haptische Guiding Methode für CVR finden, was möglicherweise dem momentanen Stand der Technik auf diesem Gebiet geschuldet ist. Die meiste Literatur zu diesem Thema konzentriert sich auf *off-screen* Guiding, da dies eine der Herausforderungen dieses Mediums darstellt. Es wird erwartet, dass Methoden aus herkömmlichen Filmen als *on-screen* Methoden funktionieren, wenn sich die Rol im Sichtfeld befindet. Es bedarf jedoch einiger Analysen, um die besten Optionen zu finden, sie in CVR umzusetzen.

Literatur	Diegesis	Sinne	Ziel	Referenz	Cue	Awareness	Freedom	Name
(Rothe und Hußmann, 2018)	dieg.	visuell	on-scr off-scr	world	direkt	subtil	freiwillig	
		audio	off-scr	world	direkt	subtil	freiwillig	
(Sheikh u. a., 2016)	dieg.	visuell	on-scr	world	direkt indirekt	subtil	freiwillig	
(Rothe und Hußmann, 2018)	n-dieg	visuell	on-scr off-scr	world	direkt	subtil apparent	freiwillig	SGD
(Nielsen u. a., 2016)	dieg.	visuell	off-scr	world	world	subtil	freiwillig	
	n-dieg	/	off-scr	screen	direkt	apparent	force-sys	
(Y.-C. Lin u. a., 2017)	n-dieg	/	off-scr	screen	direkt	apparent	force-sys	Auto-pilot
		visuell	off-scr	world	indirekt.	apparent	freiwillig	Arrow
(Gugenheimer u. a., 2016b)	n-dieg	/	off-scr	world	direkt.	apparent	force-sys	SwiVRC hair
(Danieau u. a., 2017)	n-dieg	visuell	on-scr off-scr	world	direkt.	apparent	freiwillig	Fading, Desaturation
(Chang u. a., 2018)	n-dieg	haptisch	off-scr	screen	indirekt	apparent	freiwillig	Face-Push
(Sassatelli u. a., 2018)	n-dieg	/	off-scr	screen	direkt	apparent	force-sys	Snap-Change
(Gruenefeld u. a., 2018b)	n-dieg	visuell	off-scr	screen	indirekt	apparent	freiwillig	Radial-Light

Tabelle 2-9: Veröffentlichte Guiding Methoden für CVR und die Attribute entsprechend der erarbeiteten Taxonomie („dieg.“ = „diegetisch“, „n-dieg.“ = „nicht-diegetisch“, „src“ = „screen“, „force-sys“ = „system-forciert“).

2.3.1 Guiding Methoden in framed Filmen und Bildern für CVR

In diesem Abschnitt werden bekannte Guiding Methoden vorgestellt, die in herkömmlichen Filmen oder Bildern verwendet werden. Sie werden gemäß der Taxonomie aus Abschnitt 2.2 klassifiziert. Unter Verwendung der Taxonomie konnten Unterschiede und Ähnlichkeiten von Methoden gefunden und eindeutige Merkmale identifiziert werden.

2.3.1.1 Filmische Methoden: diegetisch, visuell/auditiv, on-screen/off-screen, world-referenziert, subtil, freiwillig

Beim traditionellen Filmemachen lenken Bewegungen, Geräusche oder Lichter, die in die Geschichte eingebunden sind, die Aufmerksamkeit des Zuschauers (Bordwell und Thompson, 2013). Solche diegetischen Techniken können den Blick auch in CVR lenken, wenn die Rol sich im Sichtfeld (on-screen) befindet. Selbst wenn es sich bei den meisten diegetischen Methoden um *on-screen* Techniken handelt, gibt es einige Ausnahmen, bei denen solche Methoden auch auf das die Lenkung in Regionen außerhalb des Screens angewendet werden können (off-screen):

- *Diegetische visuelle* Cues: Wenn eine Person in eine Richtung außerhalb des Screens schaut, folgt der Betrachter oft dieser Blickrichtung (Sheikh u. a., 2016). Gleiches gilt für bewegte Objekte (Rothe und Hußmann, 2018).
- *Diegetische auditive* Cues: Töne motivieren den Benutzer, nach der Quelle des Tons zu suchen und daher die Blickrichtung zu ändern (Rothe und Hußmann, 2018).

Diegetische Cues sind *subtil* und der Betrachter kann ihnen *freiwillig* folgen. Aufgrund der Natur der diegetischen Methoden beziehen sie sich immer auf die VR/Film-Welt. Diegetische Cues sind meist direkt, jedoch sind Ausnahmen denkbar, zum Beispiel eine Person, die über ein Objekt im Raum spricht, wäre ein indirekter, diegetischer Cue.

In den **Studien 7.1.1** und **7.1.2** wurden solche diegetischen Cues von uns untersucht. Vor allem auditive Cues eignen sich gut zur Aufmerksamkeitslenkung, da sie auch im *off-screen* Fall wirken. Aber auch bewegte Elemente sind für das Guiding geeignet.

2.3.1.2 Bildmodulation: nicht-diegetisch, visuell, on-screen, world-referenziert, subtil/apparent, freiwillig

Bildmodulationen wie das Ändern der Farbe, des Glanzes oder der Sättigung sind meist nicht-diegetisch, visuell, on-screen, world-referenziert und *freiwillig*. Ob die Modulation subtil oder apparent ist, hängt vom Grad der Modifikation ab. Sowohl Saliency-Modulationen wie auch Unschärfen sind Effekte, die in herkömmlichen Filmen verwendet werden, um die Aufmerksamkeit des Betrachters zu lenken. Danieau et al. (Danieau u. a., 2017) wendeten sie auf CVR an und verglichen vier Videoeffekte für CVR: (1) Überblenden zu Schwarz für den Bereich außerhalb der RoI, (2) Entsättigung (wie SMT), (3) Unschärfe und (4) Verformung an der Seite des FoVs. In einer informellen Anwenderstudie waren Unschärfe und Deformation beim Lenken nicht erfolgreich. Es wurde ein Zusammenhang zwischen der Effizienz und der Wahrnehmbarkeit der Effekte gefunden. Die Methoden waren entweder störend (Überblenden zu Schwarz) oder unwirksam (Entsättigung). In einigen unserer **Studien 7.1.4 -7.1.6** haben wir ähnliche Erfahrungen gemacht (Rothe u. a., 2018a): Die implementierten Flicker-Methoden funktionierten nicht, solange die Flicker subtil waren. Bezüglich der Unschärfemethoden, welche von uns auch getestet wurden, vermuten wir, dass die Auflösung von CVR-Filmen und VR-Displays noch nicht hoch genug ist, um einen relevanten Unterschied zwischen dem unscharfen und dem nicht unscharfen Bereich festzustellen.

2.3.1.3 Overlays: nicht-diegetisch, visuell, on-screen/off-screen, world-referenziert/screen-referenziert, apparent, freiwillig

Overlays wie Pfeile sind *indirekte* Indikatoren. Es bedarf einer Interpretation, um die korrekte Richtung zu finden. Solche Methoden sind sowohl screen-referenziert wie auch world-referenziert anwendbar. Lin et al. (Y.-C. Lin u. a., 2017) verglichen einen Pfeil (indirekter Cue) mit einer forcierten Rotation. Beide Methoden sind apparent und wurden für ein Sportvideo und eine Stadtrundfahrt untersucht. Im Gegensatz zur forcierten Rotation benötigt der Zuschauer mit dem Pfeil mehr Zeit, zum Ziel zu gelangen und kann frei entscheiden, ob er dem Cue folgt.

2.3.1.4 Subtle Gaze Direction (SGD): nicht-diegetisch, visuell, on-screen/off-screen, world-referenziert/screen-referenziert, subtil, freiwillig

Für statische Bilder auf Bildschirmen wurden verschiedene SGD-Methoden untersucht (Bailey u. a., 2012; McNamara u. a., 2008). Solche Methoden können den Erfolg bei Suchaufgaben verbessern (McNamara u. a., 2008) und die Fehlerrate beim Merken von Regionen und ihren Positionen verringern (Bailey u. a., 2012). Um diese Methoden auf CVR oder VR auszudehnen, müssen verschiedene Aspekte berücksichtigt werden (**Studien 7.1.4 - 7.1.6**):

- Eine für statische Bilder entwickelte Methode funktioniert nicht automatisch in bewegten Bildern, in denen sich außer der Stimuli auch andere Regionen ändern.

- Eine für klare Testumgebungen entwickelte Methode funktioniert nicht automatisch bei komplexen Bildern oder Videos, in denen viele Objekte um Aufmerksamkeit konkurrieren.
- Eine für einen Monitor entwickelte Methode muss für den Fall erweitert werden, in dem sich das Zielobjekt nicht im FoV befindet.
- Bei einer Methode mit Flickern muss die Bildrate des Films und des HMD berücksichtigt werden.

In unseren SGD-Studien für CVR (**Studien 7.1.4 - 7.1.6**) (Rothe u. a., 2018a) erzielten wir ähnliche Ergebnisse wie Danieau et al. für Videoeffekte (Danieau u. a., 2017): Die Suche nach den richtigen Parametern für die Methode führte zu Techniken, die entweder nicht subtil waren oder nicht gut funktionierten. SGD, welches für Standbilder gut funktioniert, ist schwer an CVR anzupassen. Das kann an der momentan verfügbaren Hardware liegen. Je nach verwendetem SGD-Typ sind hohe Displayfrequenzen oder ein breites Sichtfeld erforderlich. Um die unterschiedliche sensorische Wahrnehmung von Fovea und Peripherie des Auges zu nutzen, scheint das FoV eines HMDs nicht groß genug zu sein. Für die Anpassung subtiler Hochfrequenz-Flicker-Methoden wird eine höhere Frequenz des HMD-Displays benötigt. Außerdem können Bewegungen im Film subtile Cues unwirksam machen. Bei nicht subtilen Flickern konnten wir in **Studie 7.1.5** jedoch eine höhere Recall-Rate feststellen.

2.3.2 Guiding Methoden in VR und AR für CVR

Im folgenden Abschnitt werden Guiding Methoden aus VR und AR diskutiert und nach der Taxonomie klassifiziert. Für jede Methode wird erläutert, ob und wie sie an CVR angepasst werden kann.

2.3.2.1 Pfeile und andere Zeichen: nicht-diegetisch, visuell, on-screen/off-screen, world-referenziert/screen-referenziert, apparent, freiwillig

Pfeile werden häufig zum Anzeigen von Richtungen im realen Leben verwendet. In mehreren Arbeiten wurden sie mit anderen Guiding Methoden verglichen (Bork u. a., 2018; Burigat u. a., 2006; Y.-C. Lin u. a., 2017; Renner und Pfeiffer, 2017b). Sie funktionieren gut, können aber das Seherlebnis beeinträchtigen. In **Studie 7.1.4** wirkte der Pfeil störender auf die CVR-Experiences als der Flicker. Der etwas dezentere Pfeil in **Studie 7.4.8**, um auf den Sprecher bei Untertiteln aufmerksam zu machen, wurde dort allerdings gegenüber der *line*-Methode bevorzugt. Für die Verwendung solcher Methoden ist der Typ der Anwendung und das Design der Zeichen von entscheidender Bedeutung.

AR-Methoden wie Attention-Funnel (Biocca u. a., 2007, 2006) oder ParaFrustum (Sukan u. a., 2014) könnte für Instruktions- oder Lernanwendungen geeignet sein. Beides wird durch augmentierte Elemente realisiert, die beim Betrachter beginnen und in die Rol führen. Die Methoden sind bemerkbar (*apparent*) und nur für eine Rol anwendbar. Da diese Overlays die Rol teilweise abdeckt, ist es für CVR-Filmerlebnisse weniger geeignet.

2.3.2.2 Stilistisches Rendering: nicht-diegetisch, visuell, on-screen, world-referenziert, subtil/apparent, freiwillig

Stilistische Rendering-Methoden (Cole u. a., 2006) für 3D-Modelle ähneln den in 2.1.4.2 beschriebenen Bildmodulationsmethoden und können an CVR angepasst werden. Das Finden des perfekten Rendering-Stils für jede Rol kann ein kreativer Teil des CVR-Designprozesses sein. Damit ein solcher Effekt wirken kann, muss er sich jedoch im Sichtfeld befinden (on-screen).

2.3.2.3 Picture-in-Picture Fenster: nicht-diegetisch, visuell, off-screen, screen-referenziert, apparent, freiwillig

Alle bisher beschriebenen Methoden geben die Richtung der Rol an. Die Darstellung der Rol in einem kleinen PiP-Fenster (Abbildung 2-4) auf dem Display bietet dagegen den Vorteil, dass der Betrachter weiß, was ihn erwartet, und somit entscheiden kann, ob er die Blickrichtung ändert. Ein Nachteil dieser Methode ist, dass das Fenster einen Teil des Inhalts abdeckt. Ein anderer Nachteil, die fehlende

Information über die Position der RoI, kann behoben werden, indem das Display auf der Seite platziert wird, die näher an der RoI ist.

Lin et al. (Y.-C. Lin u. a., 2017) untersuchten eine ähnliche Methode für sphärische Filme auf mobilen Geräten. Die *PiP*-Methode (Picture-in-Picture) lieferte dabei in den meisten Punkten bessere Ergebnisse als das pfeilbasierte Guiding, auch wenn sie mehr Platz beansprucht.



Abbildung 2-4: Zwei Guiding Methoden für CVR (Screenshots aus unserer Arbeit). Links: RoI angezeigt durch einen Pfeil. Rechts: RoI angezeigt durch ein PiP-Fenster.

2.3.2.4 Radar: nicht-diegetisch, visuell, off-screen, screen-referenziert, apparent, freiwillig

Methoden aus Kollisionsvermeidungssystemen für Segelflugzeuge könnten ebenfalls verwendet werden, um auf Rols hinzuweisen. Solche Systeme zeigen aus welcher Richtung ein anderes Segelflugzeug kommt. Abbildung 2-5 zeigt zwei Beispiele, um die Richtung der RoI anzuzeigen. Solche Methoden werden auch genutzt, um sphärische Videos in Desktop-Umgebungen zu (Chambel u. a., 2011; Neng und Chambel, 2010) und Menschen mit Höreinschränkungen bei der Sprecherfindung zu unterstützen (Agulló u. a., 2019; Montagud u. a., 2020). In **Studie 7.4.1** haben wurde die Methode in Abbildung 2-5 (links) genutzt, um anzuzeigen, in welche Richtung der andere Zuschauer bei einem gemeinsamen CVR-Erlebnis schaut.

Radar-Methoden sind screen-referenziert: Die Position des Radars ist stabil auf dem Display und ändert die Position bei Kopfdrehung nicht. Sie eignen sich vor allem dazu, um auf Regionen außerhalb des Sichtfeldes aufmerksam zu machen (off-screen).



Abbildung 2-5: Kollisionsvermeidungsmethoden von Segelflugzeugen, die übertragen wurden, um die Richtung der RoI anzuzeigen. In beiden Fällen befindet sich die Richtung auf der linken Seite hinter dem Betrachter, unterhalb der eigenen Blickrichtung. Links: Der Balken unten zeigt an, ob sich die RoI rechts oder links befindet. Der Balken rechts zeigt an, in welcher Höhe sich die RoI befindet. Rechts: Die horizontale Richtung wird durch einen Kreis und die Höhe durch einen Balken angezeigt.

2.3.3 Designüberlegungen bei der Anwendung der Taxonomie

Die eingeführte Taxonomie unterstützt Wissenschaftler und Praktiker bei der Entwicklung von Guiding Methoden für CVR. Der folgende Abschnitt verbindet die Dimensionen mit den Designfragen des CVR-Entwicklers. Bei der Umsetzung eines CVR-Erlebnisses kennen die Entwickler Material und Zielsetzung,

um sich mit der gewünschten Wirkung einer Methode und deren Eigenschaften auseinanderzusetzen. Für Entscheidungen über die geeignetste Guiding Methode sind verschiedene Aspekte relevant:

- Wie schnell soll die Methode arbeiten? (*Tempo*)
- Ist Presence wichtiger als die Wirksamkeit der Methode? (*Wirksamkeit*)
- Ist es wichtig, dass sich der Betrachter an das Ziel erinnert? (*Recall-Rate*)
- Gibt es mehrere Rols gleichzeitig? (*Anzahl der Rols*)
- Gibt es ein Problem, wenn Indikatoren (Pfeile, Halo) den Filminhalt abdecken? (*Abdeckung*)
- Wie komplex ist der Inhalt? (visueller oder auditiver *Clutter*)
- Sollte die Guiding Methode Teil des CVR-Films sein? (*Erlebnis*)

Die Beantwortung dieser Fragen ist der erste Schritt, die richtige Technik zu finden.

Tempo: In herkömmlichen Filmen kann der Filmemacher das Tempo bestimmen, indem Bildausschnitte für kurze oder lange Zeit gezeigt werden. In CVR erkundet der Benutzer die Szene durch die Änderung der Blickrichtung selbst im eigenen Tempo. Der Filmemacher kann diesen Prozess beeinflussen, indem er die richtige Guiding Methode wählt. Um das Tempo eines Films zu beeinflussen, kann man zwischen langsam und schnell wirkenden Methoden wählen. Zum Beispiel funktionieren forcierte Methoden sehr schnell, können aber die Experience zerstören. Es sind jedoch auch CVR-Filme denkbar, bei denen eine forcierte Rotation Teil der Experience ist.

Wirksamkeit: Es hängt vom Zweck des Films ab, ob die Wirksamkeit einer Methode wichtiger ist als Presence. Apparente Techniken sind effektiver, können aber das Filmerlebnis zerstören. Für manche Stories ist es möglicherweise weniger bedeutend, dass die Zuschauer immer den Cues folgen. In diesem Fall sollten subtile, freiwillige Methoden bevorzugt werden. Im Gegensatz dazu kann es für ein Sportereignis wichtig sein, die Details zum richtigen Zeitpunkt zu sehen. Hier können forcierte oder stimulusgesteuerte Methoden eine gute Wahl sein. Bei Lehrfilmen passt ein Pfeil. Es ist eine deutlich wahrnehmbare, apparente Technik, die einen indirekten Cue verwendet, der schwer zu übersehen ist.

Recall Rate: Wenn der CVR-Film als Lernanwendung gedacht ist, kann die Recall-Rate wichtig sein, und es können offensichtlichere Methoden angewendet werden, beispielsweise apparente, nicht-diegetische. Auch für andere Genres kann es relevant sein, dass sich der Zuschauer Details der Geschichte merkt, der Indikator sollte jedoch nicht zu präsent sein. Basierend auf der Aufmerksamkeitstheorie sind freiwillige Prozesse gedächtnisgebunden, halten die Aufmerksamkeit länger und erhöhen so die Recall-Rate. Auch Modulationstechniken (z. B. Stylistic Rendering, SMT) sind geeignet, da sie die Merkfähigkeit beeinflussen (Veas u. a., 2011).

Anzahl der Rols: Die oben genannten Methoden werden meist für eine einzelne Rol untersucht. Insbesondere für nichtlineares Storytelling kann aber mehr als eine Rol gleichzeitig erforderlich sein. Nicht alle der genannten Methoden sind in der Lage, dies zu unterstützen. Auch wenn einige Methoden für mehr als eine Rol gleichzeitig angewendet werden können, kann dies zu einer Überfüllung der Anzeige und zu einer Überlastung des Betrachters führen. Es sind weitere Untersuchungen erforderlich, um Methoden und Anpassungen zu finden, mit denen mehrere ROIs gleichzeitig angezeigt werden können.

Abdeckung: Diegetische oder Modulationsmethoden decken keine Filminhalte ab. Bei Verwendung von Indikatoren wie Pfeilen oder Fenstern sind Teile des Films nicht sichtbar. Dies kann störend auf den Zuschauer wirken. Es sind jedoch Anwendungen denkbar, bei denen der Indikator wichtiger ist als die vollständige Sichtbarkeit, z. B. Anleitungsvideos. Die Auswahl der richtigen Parameter (Größe oder Farbe) kann den Cue offensichtlicher/effektiver oder subtiler machen. World-referenzierte Overlays (z. B. Pfeile) bleiben an derselben Stelle im Film und decken einen Bereich dauerhaft ab. Screen-referenzierte Overlays (z. B. Zeichen am Displayrand) ändern den abgedeckten Bereich, wenn sich der Kopf bewegt.

Clutter: Um beurteilen zu können, ob eine Methode geeignet ist, muss die Komplexität eines Bildes berücksichtigt werden. Überfüllte Bilder mit vielen Details erfordern klare und offensichtliche Techniken. Wenn das Bild klar ist, können subtilere Methoden funktionieren. Gleiches gilt für Audio: Wenn der Film

viele Geräusche enthält, kann es schwierig sein, einem räumlichen Audiosignal zu folgen, das zu einer Rol führen sollte.

Erlebnis: Nicht immer ist es notwendig, das Guiding subtil zu gestalten. Es kann ebenso Teil des Erlebnisses sein, ähnlich wie Szenenübergänge den Film beeinflussen. Die Guiding Methode kann den Stil, das Tempo und die Atmosphäre eines Films prägen.

Insgesamt bietet die obige Taxonomie Unterstützung bei der Beantwortung dieser praktischen Fragen und bei der Suche nach der geeignetsten Technik.

2.4 Fazit und Ausblick

Basierend auf vorhandener Literatur und eigenen Studien wurde eine Taxonomie für Guiding Methoden beschrieben. Auch wenn dabei der Fokus auf CVR-Anwendungen lag, sind die meisten Begriffe und Attribute auch auf AR und VR übertragbar. Wir haben die Vor- und Nachteile von Attributen aufgelistet und jede Dimension anhand konkreter Beispiele für Guiding Methoden veranschaulicht. Diese Taxonomie kann dazu beitragen, die verschiedenen Merkmale von Guiding Methoden zu verstehen und unterstützt Wissenschaftler und Praktiker dabei, die richtige Technik für unterschiedliche Anforderungen zu finden.

Die Anwendung von Indikatoren zur Aufmerksamkeitslenkung ist ein zusätzliches Mittel für Filmemacher, um Stil und Tempo des Films auf neuartige Weise zu prägen. Ähnlich wie Szenenübergänge in framed Filmen müssen die passenden Guiding Methode für jede Rol gesucht werden. Ein Wechseln zwischen den Techniken in einem Film kann nützlich sein, um z. B. das Tempo bestimmter Filmabschnitte zu ändern.

Weitere Forschungen sind notwendig, um neue Wege zu finden, wie Guiding Methoden in Stories integrierbar sind. Dabei ist es insbesondere notwendig, diegetische und subtile Methoden weiter zu untersuchen. Eine große Rolle spielt dabei der Ton, da dieser auch wahrnehmbar ist, wenn die Tonquelle nicht im aktuellen FoV ist. Aber auch haptische Elemente erfordern weitere Untersuchungen.

3

How to Shoot: Kameraführung in CVR

„Jede Sprache basiert auf einer allgemein akzeptierten Konvention. Eine Sprachgemeinschaft einigt sich darauf oder lernt, bestimmten Symbolen bestimmte, für alle Mitglieder verbindliche Bedeutungen zuzuordnen. Wer eine Geschichte erzählen oder Ideen verbreiten will, muss zuerst diese Symbole lernen und die Regeln, nach denen sie miteinander kombiniert werden. Doch all dies - Symbole, Bedeutungen, Regeln - befindet sich in ständiger Weiterentwicklung. Künstler oder Philosophen können die Sprachgemeinschaft beeinflussen, indem sie neue Symbole oder Regeln einführen und alte ausmustern. Dem Kino ist dieser Prozess nicht fremd.“

Daniel Arijon (Arijon, 2000)

Kamerahöhen, -abstände und -bewegungen sind wichtige Stilelemente im Film. Da der Zuschauer in der virtuellen Welt die Position der Kamera einnimmt, müssen einige der traditionellen Elemente der Kameraführung für CVR neu überdacht werden. Die Augenhöhe in der realen Welt variiert durch unterschiedliche Körpergrößen und -haltungen, wodurch die Wahrnehmung der *Kamerahöhe* in der virtuellen Welt beeinflusst wird. Ein zweiter Aspekt ist die Auswahl des *Kameraabstandes*, welcher Gefühle von Nähe oder Distanz entstehen lassen kann. Schließlich muss auch der Einsatz von *Kamerabewegungen* hinterfragt werden, da eine solche Bewegung in der virtuellen Welt nicht konform ist mit den sensorischen Bewegungsinformationen der fixen Position in der realen Welt. In diesem Kapitel werden diese Herausforderungen diskutiert und Guidelines für die Kameraführung in CVR entwickelt.

Dieses Kapitel enthält bereits veröffentlichtes Material aus folgenden Publikationen:

- Rothe, S., Kegeles, B., Hussmann, H., 2019. Camera Heights in Cinematic Virtual Reality. in: Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '19
- Rothe, S., Kegeles, B., Allary, M., Hussmann, H. 2018. The Impact of Camera Height in Cinematic Virtual Reality. in: Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology - VRST'18
- Rothe, S., Sarakiotis, V., Hussman, H., 2019. Where to Place the Camera: How Viewers React to Different Camera Positions in Cinematic Virtual Reality. in: Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology - VRST'19

In CVR gibt es zwei wesentliche Herausforderungen für die Kameraführung. Zum einen befindet sich der Zuschauer während der CVR-Experience an der Stelle der Kamera während der Aufnahme. Zum anderen ist stets nur ein Teil des sphärischen Bildes sichtbar. Dieser wird vom Zuschauer selbst bestimmt und der CVR-Entwickler hat auf die Auswahl des Bildausschnittes nur begrenzte Einflussmöglichkeiten.

Ähnlich wie in framed Filmen kann in CVR die Kamerahöhe für die Komposition von Szenen verwendet werden. Durch die Kamerahöhe und damit die Perspektive der Kamera lassen sich Gefühle wie Erhabenheit oder Bedeutungslosigkeit transportieren. In vielen Situationen ist eine neutrale Position erwünscht. Diese neutrale Position ist aufgrund unterschiedlicher Körpergrößen und -haltungen in CVR schwer realisierbar. Nur selten ist es der Fall, dass die Kamerahöhe bei der Aufnahme (virtuelle Augenhöhe) der Augenhöhe des Zuschauers (reale Augenhöhe) entspricht. In Abschnitt 3.2 wird untersucht, welche Auswirkungen Diskrepanzen zwischen diesen beiden Werten auf das Zuschauererlebnis haben.

Eine wichtige Aufgabe der Kameraführung ist die Festlegung des richtigen Bildausschnittes - die *Kadrierung*. Auf diese Weise wird unter anderem die passende Einstellungsgröße gewählt, um eine Person oder ein Objekt darzustellen. Dabei spielen sowohl die Komposition des Bildes wie auch der Abstand der Kamera eine Rolle. Die Auswahl des Bildausschnittes liegt in CVR allerdings nicht mehr allein in der Verantwortung des Filmemachers, sondern wird vom Betrachter beeinflusst. Wie Einstellungsgrößen in Kameraentfernungen umgesetzt werden können und wie sich unterschiedliche Anordnungen der Charaktere auf das Zuschauererlebnis auswirken, wird in Abschnitt 3.3 untersucht.

Um den Zuschauer zu verschiedenen Orten bringen zu können, sind Schnitte und auch Kamerabewegungen nutzbar. Diese beiden Methoden werden in Abschnitt 3.4 miteinander verglichen und die Vor- und Nachteile erörtert.

Um CVR-Entwicklern Hilfestellungen für die Kameraarbeit zu geben, werden in Abschnitt 3.5 die Ergebnisse der vorhergehenden Abschnitte in Guidelines zusammengefasst.

3.1 Theoretischer Hintergrund

3.1.1 Regeln der Kameraführung

3.1.1.1 Mis-en-scène

In einem Film beeinflusst die Anordnung von Personen, Objekten und Kamera, wie Objekte/Personen und ihre Verbindungen zueinander wahrgenommen werden (Arijon, 1991). Dieser räumliche Grundaufbau wird als *Mis-en-scène* (engl.: placing on stage) bezeichnet. Der Begriff wurde Mitte des 20. Jahrhunderts von André Bazin aus dem Theater in die Filmästhetik übernommen (Wulff, 2012c). Ursprünglich

entstanden, um dem etablierten Prinzip des Filmschnittes (*Montage*) in den 1950er etwas entgegenzusetzen, ergänzen sich heutzutage beide Prinzipien. Bazin beschreibt *Mis-en-scène* als Stilmittel und setzt auf einen gut überlegten sorgfältigen Bildaufbau. Dabei wird eine Szene nicht in mehrere Einstellungen aufgelöst, sondern durch Schärfeverlagerungen, Lichtgestaltung, Kamerabewegungen und Farbkompositionen zu einer flüssigen Handlung gestaltet. Solche Sequenzen, die nur aus einer einzigen Einstellung bestehen, werden *Plansequenzen* genannt und stellen hohe Anforderungen an das filmische Können. Mehrere Filmemacher stellten sich der Herausforderung und drehten ganze Filme mit Plansequenzen, z. B. (Hitchcock, 1948) „Rope“ und (Sokurow, 2002) „Russian Ark“.

Bazin vertrat die Meinung, dass durch die Montage, bei der eine Szene in Einstellungen mit verschiedenen Einstellungsgrößen und Blickwinkeln fragmentiert wird, ein Film weniger realistisch ist und die Zuschauer weniger involviert werden. Auch wenn diese Behauptung von Buccalo widerlegt wurde (Buccalo, 1977), ist *Mis-en-scène* ein wichtiges Gestaltungsmittel, welches in CVR an Bedeutung gewinnt, da durch das vollständige Rundumbild und die zuschauerinitiierte Bildauswahl Schnitte eine andere Bedeutung erhalten.

Da sich der Betrachter in CVR als Teil der Szene fühlt, ist es noch wichtiger als in framed Filmen, wie Personen und Objekte zueinander und zum Betrachter (Kamera) angeordnet sind. Zu nahestehende Charaktere könnten z. B. Unbehagen oder sogar Angstzustände hervorrufen.

3.1.1.2 Kameraperspektiven

Die Kameraperspektive beschreibt den Blickwinkel der Kamera auf ein Motiv und wird durch die Höhe der Kamera bestimmt (Abbildung 3-1). Eine *Untersicht*-Kamera schaut nach oben und macht den Charakter größer und je nach Kontext selbstbewusst, heldenhaft oder auch bedrohlich, arrogant (Kandorfer, 2003). Eine *Obersicht*-Kamera schaut auf Objekte oder Charaktere herab und diese erscheinen kleiner, einsamer oder schutzbedürftiger (Kandorfer, 2003). Der Normalfall, die Kamera auf Augenhöhe zu platzieren, ist in CVR aufgrund der unterschiedlichen Körpergrößen nicht ohne Weiteres zu realisieren. Wenn eine kleinere Person eine Person auf Augenhöhe anschaut, die größer oder kleiner aussieht als sie selbst, können Diskrepanzen in der Wahrnehmung verursacht werden.

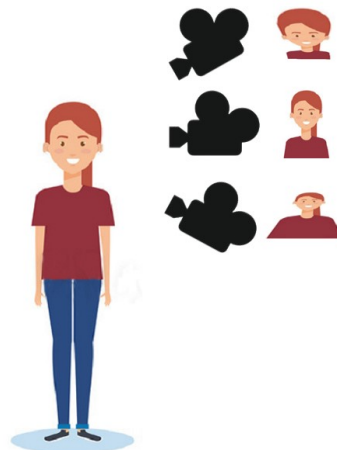


Abbildung 3-1: Kameraperspektiven: Übersicht-Kamera, Kamera auf Augenhöhe, Untersicht-Kamera.

Außerdem können unterschiedliche Kamerawinkel genutzt werden, um die Erzählperspektive zu wechseln, einen Schauplatz vorzustellen oder Objekte zu verfolgen (Katz, 2000).

3.1.1.3 Positionen und Einstellungsgrößen

Einstellungsgrößen

Beim traditionellen Filmemachen sind verschiedene Einstellungsgrößen ein wichtiges Werkzeug für das filmische Erzählen und können zur Vermittlung emotionaler Informationen verwendet werden (Bordwell und Thompson, 2013). Es gibt verschiedene Begriffe und Schemata zum Definieren von Einstellungsgrößen, wobei die Grenzen zwischen den Größen nicht einheitlich sind und es auch kulturelle Unterschiede gibt. Extrem weite Einstellungen (Supertotale) geben einen Überblick über die Szenerie und werden oft verwendet, um in einen Film einzuführen. Nahaufnahmen (Close-up) einer Person zeigen deren Gefühle, und der Zuschauer fühlt sich ihr nah. Die Einstellungsgrößen werden durch den Kameraabstand und Parameter der Kamera (z.B. eingestellte Brennweite) beeinflusst. Abbildung 3-2 und Abbildung 3-3 zeigen die wichtigsten Einstellungsgrößen für die Darstellung einer Person.

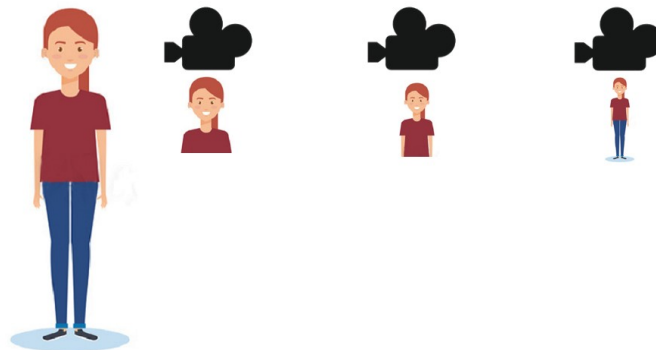


Abbildung 3-2: Kameraabstände und Einstellungsgrößen. Von links nach rechts: Naheinstellung, Halbnahe und Totale.

Die Wahl verschiedener Einstellungsgrößen ist ein Stilmerkmal des Films. Unterschiedliche Filmmacher, Filmtraditionen und Kulturen setzen sie in unterschiedlicher Weise ein. Da der Betrachter in CVR das Sichtfeld frei wählen kann, ist es schwieriger, Einstellungsgrößen zu verwenden. Die Kadrierung wird nicht vom Filmmacher festgelegt, und ein Wechsel zwischen den Kameraentfernungen erzeugt beim Betrachter das Gefühl zu springen.

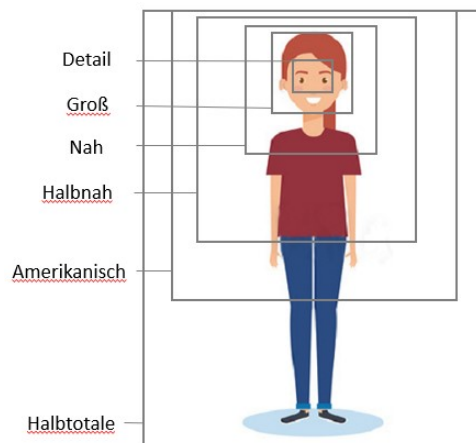


Abbildung 3-3: Verschiedene Einstellungsgrößen für eine Person.

Zusätzlich zur Kameraentfernung kann bei framed Filmen das Ändern der Brennweite die Einstellungsgröße beeinflussen. Mit einer großen Brennweite ist der aufgenommene Ausschnitt der Umgebung kleiner, die Objekte wirken näher und die Entfernungen kürzer. Aufgrund der Tatsache, dass in CVR das sphärische Bild mit einer vollständigen Rundumsicht gefüllt ist, können Änderungen der Brennweite nicht ohne Weiteres zur Änderung der Einstellungsgröße genutzt werden.

Abstände zwischen Kamera und Darsteller

Da der Betrachter in CVR in der Szene die Position der Kamera einnimmt, wird die Entfernung zwischen Kamera und Charakter während der Aufnahme zur Entfernung zwischen Betrachter und Charakter in der virtuellen Welt. In den Studien von Sheikh et al. (Sheikh u. a., 2016) bevorzugten die Teilnehmer Entfernungen zwischen 2 und 3 m, um sich ein Kampftraining anzusehen. Für einige Teilnehmer waren 2 m jedoch zu nah, sie hatten das Gefühl, ihr persönlicher Raum wäre verletzt. Eine Entfernung von 4 m wurde als zu weit entfernt empfunden. Die Fähigkeit, der Geschichte zu folgen, wurde von der Entfernung nicht beeinflusst.

Keskinen et al. vergleichen zwei Kameraentfernungen zum Charakter in einer CVR-Umgebung: entfernt (zwischen 3 und 4 m), nahe (ungefähr 1 m) mit entfernungsändernden Charakteren. Sie schlagen einen Abstand zwischen Kamera und Charakter von mehr als 1 m vor.

Abstände zwischen Darstellern

Abstände zwischen Charakteren können nicht einfach aus der realen Welt übernommen werden. Da das FoV von HMDs kleiner ist als unser übliches FoV (ungefähr 180° horizontal), können in einigen Szenarien Charaktere in der virtuellen Welt nicht gleichzeitig gesehen werden, selbst wenn dies in der realen Welt mit den gleichen Positionen möglich wäre. Dies wird in **Studie 7.2.2** näher untersucht.

F-Formationen

Neben den Entfernungen ist die Anordnung der Charaktere zueinander wichtig. Solche Anordnungen, die als *F-Formationen* (facing formations) bezeichnet werden, sind Gegenstand soziologischer Forschung und werden vom Typ der Aufgabe beeinflusst, die Menschen gemeinsam erledigen, z. B. kompetitive, kollaborative oder kommunikative Aufgaben (Kendon, 2010; Marshall u. a., 2011). Abbildung 3-4 zeigt einige Beispiele solcher F-Formationen für zwei Personen.

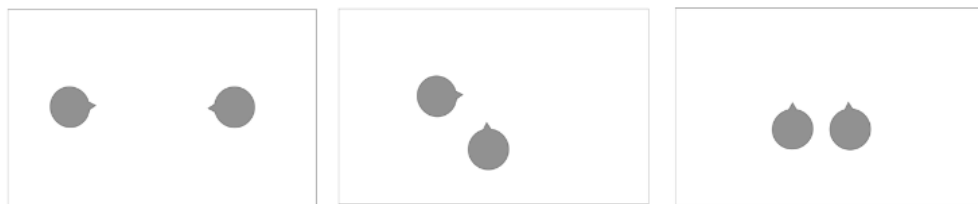


Abbildung 3-4: Beispiele für F-Formationen: links: gegenüber (z. B. Gespräch), Mitte: L-Formation (z. B. gemeinsames Arbeiten), rechts: nebeneinander (z. B. gemeinsames Filmschauen).

In CVR ist die Nutzung des Raumes für die Geometrie des Storytellings wichtig (Pope u. a., 2017). Ähnlich wie im Theater wird dem Zuschauer die gesamte Szenerie und nicht wie in Filmen ein vorausgewählter Bildausschnitt angeboten. Pope et al. (Pope u. a., 2017) verglichen, wie Theaterpraktiker eine Theaterszene mit räumlichen Interaktionen auf eine Szene für einen CVR-Film übertragen. Für die Theater-Szene sollten die Teilnehmer diese Szenerie um einen auf einem Drehstuhl sitzenden Zuschauer arrangieren, für die CVR-Szene um die Kamera. Es gab mehrere Unterschiede zwischen den beiden Ergebnissen: In der Theaterszene wurde der Zuschauer häufiger in den Mittelpunkt der Interaktion gestellt. Außerdem waren die Abstände zwischen den Charakteren in CVR kürzer, ebenso wie die Abstände zwischen Betrachter und Charakteren.

3.1.1.4 Kamerabewegungen

Kamerabewegungen können für verschiedene Zwecke verwendet werden, z. B. um die Aufmerksamkeit des Betrachters zu lenken, um den Zuschauer zu einem neuen Ort zu führen oder visuelle Effekte zu erzeugen (Katz, 2000). Oft sind Kamerabewegungen Teil der Geschichte: beispielsweise ein umherschweifender Blick oder ein fahrender Zug.

Es gibt zwei Hauptarten von Kamerabewegungen: Änderung der Position einer Kamera (Fahren) und Drehung der Kamera (Schwenk). Kamerafahrten können sowohl seitwärts (*truck*), vor und zurück (*dolly*) als auch vertikal (*pedestal*) ausgeführt werden. Auch Schwenks sind sowohl vertikal (*tilt*) wie auch

horizontal (*pan*) möglich. Bei Kameratechniken wie Zoom (*zoom*) und Schärfenverlagerung (*rack-focus*) wird die Kamera nicht als Ganzes bewegt, sondern Teile von ihr (Objektiv). Sie werden meist auch zu den Kamerabewegungen gezählt.

3.1.1.5 180°- und 30°-Regel

Schon beim Dreh müssen Regeln beachtet werden, um Aufnahmen beim Schnitt problemlos aneinanderfügen zu können. Deshalb werden die 180°- und 30°-Regel der Vollständigkeit halber hier erwähnt. Da beide Regeln vor allem für den Schnitt relevant sind, werden sie ausführlicher in Kapitel 4 behandelt. Die 180°-Regel legt nahe, Sprünge über die Handlungsachse zu vermeiden. Wohingegen die 30° Regel empfiehlt, die Kamera zwischen zwei Einstellungen um mindestens 30° zu versetzen (Beller und Wulff, 2012).

3.1.2 Proxemik

Die *Proxemik* ist ein Teilgebiet der Psychologie und der Kommunikationswissenschaften, welches untersucht, wie Menschen den Raum nutzen und wie sich Entfernungen auf Verhalten, Kommunikation und soziale Interaktion auswirken (Hall, 1966; Marquardt u. a., 2012; Setti u. a., 2014). Proxemik gibt Auskunft darüber, wie wir Objekte, Personen und uns selbst im Raum anordnen, welche Positionen intuitiv eingenommen werden und welche Entfernungen unangenehm sind.

Der Begriff *Proxemik* wurde in den 1960er-Jahren von Edward T. Hall geprägt und beschreibt den sozialen und persönlichen Raum und deren Wahrnehmung durch den Menschen (Hall, 1966). Hall war Anthropologe und Ethnologe und seine Werke über verschiedene Kulturdimensionen machten ihn zu einem der Begründer der interkulturellen Kommunikationswissenschaft. Liest man heute in Halls Werken, wird einem klar, dass nicht nur unterschiedliche Kulturen das Verhalten beeinflussen, sondern auch Zeitepochen und Generationen. Die von ihm beschriebenen Dimensionen sind: *Proxemik*, *Kontextbezug*, *Zeitverständnis* und *Informationsgeschwindigkeit*. Auch wenn es alle vier Dimensionen wert sind, auf ihre Anwendung in CVR untersucht zu werden, wollen wir uns in dieser Arbeit auf die Proxemik beschränken.

Reale Welt

Der Raum um einen Menschen wurde von Hall in vier Zonen strukturiert (Hall, 1966): (1) intimer Raum, (2) persönlicher Raum, (3) sozialer Raum und (4) öffentlicher Raum. Die Größen dieser Zonen unterscheiden sich zwischen Kulturen und Persönlichkeiten. Referenzwerte sind in Abbildung 3-5 dargestellt. Jede dieser Regionen hat ihre eigene Funktionalität:

- intime Zone: Berühren oder Flüstern
- persönliche Zone: Interaktionen zwischen Freunden oder der Familie
- soziale Zone: für die Interaktion unter Bekannten
- öffentliche Zone: öffentliche Rede

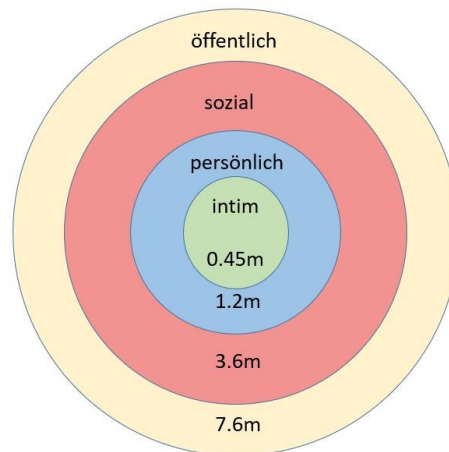


Abbildung 3-5: Entfernungen für die verschiedenen von Hall beschriebenen Zonen (Hall, 1966).

Virtuelle Welt

Es gibt mehrere Studien über die Wirkung von Proxemik in virtuellen Umgebungen und ob Menschen einen ähnlichen Abstand zu Avataren benötigen wie zu Menschen in der realen Welt: Wilcox et al. (Wilcox u. a., 2006) untersuchten den persönlichen Raum in Bezug auf die Entfernung zu virtuellen Charakteren, jedoch nicht in einer immersiven Umgebung: Bilder von realen Personen wurden auf eine Leinwand projiziert, so dass die Position während der gesamten Studie unverändert blieb. Die Teilnehmer zeigten ähnliche Reaktionen auf Verletzungen des zwischenmenschlichen Raumes wie in natürlichen Umgebungen. Auch andere Studien haben gezeigt, dass Menschen in VR auf ähnliche Weise reagieren, wie es in der Proxemik für die reale Welt beschrieben wird (Wilcox u. a., 2006) (Llobera u. a., 2010). In den Studien von Bailenson et al. (Bailenson u. a., 2001) hielten die Teilnehmer mehr Abstand zu Charakteren als zu ähnlich großen leblosen Objekten.

Die hier aufgeführten Erkenntnisse werden in Abschnitt 3.2 genutzt, um für verschiedene Zwecke in CVR die richtige Kameraposition zu finden.

3.1.3 Sichtfeld - Blickfeld - Gesichtsfeld

Die Begriffe Sichtfeld, Blickfeld, Gesichtsfeld beschreiben alle den Bereich, den eine Person sehen kann, unterscheiden sich aber in Details. Sie werden in der Literatur nicht einheitlich verwendet. Wir benutzen in dieser Arbeit die folgenden Definitionen:

Das *Blickfeld* (engl.: field of regard) ist der Bereich, den eine Person von einer Position aus sehen kann, einschließlich der Kopf- und Augenbewegungen.

Das *Sichtfeld* (engl.: field of view) ist definiert als der Bereich, den eine Person ohne Kopfbewegung mit den Augen oder durch ein optisches Gerät sehen kann (iTechWearer, 2017). Es wird horizontal und vertikal in Grad gemessen, manchmal auch diagonal. Der Begriff *Sichtfeld* wird manchmal - vor allem in technischen Spezifikationen - auf optische Geräte eingeschränkt (Wikipedia, 2018). Wir nutzen in dieser Arbeit allerdings die obige Definition, die auch das Sichtfeld des Menschen einschließt (Rouse, 2017).

Das *Gesichtsfeld* (engl.: field of vision) ist ein Begriff der Physiologie und Augenheilkunde und beschreibt den Bereich, der ohne Kopf- und Augenbewegungen sichtbar ist. Er ist aus medizinischer Sicht interessant, spielt aber in dieser Arbeit keine weitere Rolle. Es ist hier der Vollständigkeit halber erwähnt, um die Begriffe voneinander abzugrenzen und Begriffskonflikte zu vermeiden.

Reale Welt

Das durchschnittliche horizontale *FoV* des Menschen beträgt ungefähr 180° (Dreyfuss, 1967). Es unterscheidet sich zwischen Individuen und nimmt mit dem Alter ab (Ball u. a., 1988; Sekuler und Bennett, 2000).

Die Bereiche, die fähig sind, bestimmte Aufgaben zu erfüllen, sind unterschiedlich groß: präzises Sehen (2°), Worterkennung (20°), normales Sehen / Symbolerkennung (60°), Farbunterscheidung (120°) (Dreyfuss, 1967) (Abbildung 3-6).

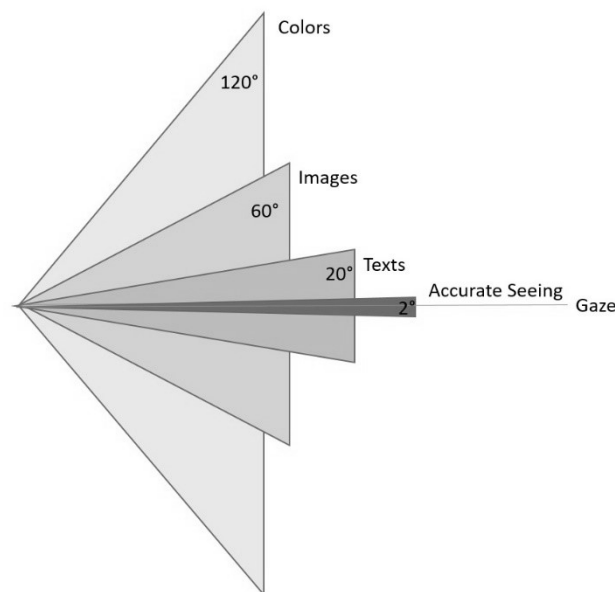


Abbildung 3-6: Gesichtsfeldbereiche des Auges, die unterschiedliche Aufgaben erfüllen können (Dreyfuss, 1967).

Virtuelle Welt

Das *FoV* ist ein wichtiges Merkmal bei der Spezifikation eines HMD und gibt die Größe des Bereichs an, der für den Betrachter sichtbar ist. Dieser ist nicht immer gleich groß und hängt vom verwendeten HMD-Typ und dem Abstand zwischen Augen und Display ab. In (Hunt, 2016) können die folgenden Werte gefunden werden:

- HTC Vive, wenn die Augen etwa 10 mm von den Objektiven entfernt sind: diagonal 145°, 100° horizontal und 110° vertikal
- Oculus Rift CV1 (Consumer-Version 1), Augen ca. 10 mm von den Objektiven entfernt: diagonal 120°, 80° horizontal und 90° vertikal
- Samsung Gear VR: zwischen 90° und 96°, abhängig vom verwendeten Smartphone, 60° horizontal, 76° vertikal

Die neuere Version der Samsung Gear VR (SM-R323) hat laut Hersteller ein *FoV* von 101° (Samsung, 2016). In unseren Studien kamen beide HMD-Versionen zum Einsatz.

Der tatsächlich wahrgenommene Teil hängt jedoch nicht nur vom HMD ab, sondern auch von anderen Faktoren, wie dem Abstand der Augen von den Linsen und der binokularen Beziehung zwischen den beiden Augen (Valve Corporation, 2017). Es kann also vorkommen, dass dem Anwender weitaus weniger als das theoretische *FoV* zur Verfügung steht.

Die verschiedenen *FoVs* beeinflussen die Größe des sichtbaren Bildausschnitts und auf diese Weise die Einstellungsgröße. Da der Filmemacher unterschiedliche *FoVs* voraussetzen muss, ist es schwierig, perfekte Kameraabstände zu finden, die in allen Umgebungen funktionieren.

3.1.4 Augenhöhe

Reale Welt

Im realen Leben nutzen Menschen Augenhöhen zur Bestimmung von Entfernungen, Objektabmessungen und zum Wahrnehmen von Affordances (Sinai u. a., 1998; Maryjane Wraga, 1999). Wraga hat gezeigt, dass die Augenhöhe mehr Einfluss auf die Bestimmung der Höhe von Objekten hat als auf die Bestimmung der Breite. Ein Grund dafür ist, dass beide, sowohl die Höhe des Objekts als auch die Augenhöhe, vertikal ausgerichtet sind (M Wraga, 1999). Dieses Wissen ist wichtig für CVR, da Höhen bekannter Objekte den Betrachter in Fällen irritieren können, in denen die Kamerahöhe nicht der physischen Augenhöhe entspricht.

Virtuelle Welt

Forscher in Medizin und Psychologie nutzen virtuelle Umgebungen, um den Einfluss der Körpergröße auf soziales Verhalten, Lebenszufriedenheit und Krankheiten wie Paranoia zu untersuchen (Freeman u. a., 2014; Leyrer u. a., 2015), (Yee u. a., 2009).

In VR ist es möglich, die virtuelle Augenhöhe einer Person zu ändern. Die Echtzeitsteuerung der virtuellen Kamera kann für interaktive Erzählungen in virtuellen Umgebungen verwendet werden (Amerson u. a., 2005; Galvane u. a., 2014; Jhala und Young, 2005). Die Größe eines Charakters relativ zum Benutzer beeinflusst, wie ein Charakter wahrgenommen wird (Marklund, 2017): größere Charaktere wirken dominanter und stärker, kleinere sozialer.

In der virtuellen Realität betrachtet der Benutzer die Umgebung über die virtuelle Kamera, jedoch entspricht die virtuelle Augenhöhe nicht immer der realen Augenhöhe des Benutzers. Es gibt verschiedene Untersuchungen, ob die virtuelle Augenhöhe die Wahrnehmung von Größen und Entfernungen in der virtuellen Welt beeinflusst (Leyrer u. a., 2011; Loomis und Knapp, 2003). Leyrer et al. (Leyrer u. a., 2015) untersuchten, wie Benutzer ihre eigene Augenhöhe in einer virtuellen Umgebung bestimmen. Beim Vergleich von visuellen und posturalen Hinweisen in virtuellen Umgebungen stellten sie fest, dass Menschen bei der Bestimmung ihrer Augenhöhe eher auf posturale Hinweise angewiesen sind, wenn ein Konflikt zwischen visuellen und posturalen Informationen besteht. In Experimenten wurde gezeigt, dass eine zunehmende Augenhöhe zu einer Abnahme der wahrgenommenen Distanz führte. Eine Verringerung der Augenhöhe hatte jedoch keinen Einfluss auf die wahrgenommene Entfernung (Leyrer u. a., 2011).

Das Steuern der virtuellen Körpergröße kann das Unbehagen verringern, welches durch unangemessene zwischenmenschliche Abstände verursacht wird. Maeda et al. (Maeda u. a., 2016) änderten virtuelle Körpergrößen, was zu unterschiedlichen Wahrnehmungen der zwischenmenschlichen Distanzen führte.

Es gibt wenig Forschungsergebnisse zu Augenhöhen in CVR. Philpot et al. (Philpot u. a., 2017) verglichen das Anschauen eines CVR-Videos mittels HMDs und CAVE-ähnlichen Displays. War die Kamera höher als die Augenhöhe, erwähnten einige Benutzer, dass sie sich groß fühlten. Eine Kamera unter Augenhöhe fühlte sich klein an. Diese Feststellung steht im Einklang mit dem Wissen über framed Filme. Bei der hohen Kamera verspürten jedoch einige Menschen ein Schwindelgefühl, was bei der niedrigen Kamera nicht der Fall war. In den Studien von Passmore et al. (Passmore u. a., 2017) beklagten sich Zuschauer, wenn die Kamera zu hoch war (etwa 1 m höher als die normale Körpergröße), einigen war schwindelig. Außerdem mussten die Teilnehmer in diesem Fall nach unten schauen und erkannten, dass sie in dieser Szene keinen Körper hatten, was kein Problem war, wenn sich die Kamera auf Augenhöhe befand. Für eine Einstellung mit einer sehr niedrigen Kamera (60 cm), die durch die Geschichte begründet wurde, äußerten einige Benutzer, dass sie sich zu klein fühlten, jedoch ohne Schwindelgefühl. In Abschnitt 3.2 interessieren uns kleinere Abstände zwischen Kamerahöhe und physischer Augenhöhe: Unterschiede, die sich aus unterschiedlichen Körpergrößen ergeben und nicht durch die Storyline motiviert sind. Wir wollen herausfinden, welche Höhen geeignet sind, ohne den Betrachter zu irritieren.

3.2 Kamerahöhe in CVR

Da der Inhalt in CVR nicht in Echtzeit erstellt wird, sondern im Voraus aufgezeichnet wird, können weder die Kamerahöhe noch die Charaktergröße für den Betrachter angepasst werden. Der Filmemacher muss die Höhe der Kamera während der Aufnahmen bestimmen. Diese Höhe sollte unterschiedliche Körpergrößen und Sitz-/Stehpositionen der Zuschauer berücksichtigen. Um dies mit nur einer Kamerahöhe zu erreichen, muss herausgefunden werden, welchen Einfluss unterschiedliche Kamerahöhen auf den Zuschauer haben. Dabei ist neben der Körpergröße auch relevant, ob der Film sitzend oder stehend angeschaut wird. Deshalb werden unterschiedliche *Postures* (Körperhaltungen) in die Untersuchungen einbezogen.

In framed Filmen ist der Betrachter in seiner realen Umgebung, und die Differenz zwischen Augen- und Kamerahöhe ist kein Problem. Das Betrachten eines Films aus der Perspektive der Kamera durch das HMD kann jedoch einige Schwierigkeiten verursachen, z. B. kann der intuitive Vergleich mit Höhen bekannter Objekte den Betrachter irritieren, falls die Kamerahöhe nicht der physischen Augenhöhe entspricht. Ziel dieses Kapitels ist es zu untersuchen, wie die Position der Kamera Presence, Simulator-Sickness und das Benutzererlebnis des Betrachters beeinflusst.

Die Höhe der Kamera spielt eine wichtige Rolle bei der Wahrnehmung von Szenen in framed Filmen. Eine Übersicht-Kamera blickt auf ein Objekt oder eine Figur und der Betrachter fühlt sich machtvoller. Eine Untersicht-Kamera schaut nach oben und macht den Betrachter kleiner. Es gibt jedoch viele Situationen in einer Geschichte, in denen der Betrachter eine neutrale Perspektive einnehmen soll. Dabei ist die Augenhöhe eines Menschen entscheidend für die menschliche Wahrnehmung. Sie wird zum Skalieren von Größen, Geschwindigkeiten und Abständen verwendet (Leyrer u. a., 2015).

Da Entfernungen und Höhen von betrachteten Objekten anhand der eigenen Augenhöhe bestimmt werden, kann eine falsche Kameraposition zu einer Desorientierung in der Filmwelt führen. In CVR-Videorichtlinien wird empfohlen, die Kamera auf Kopfhöhe zu platzieren (Jaunt, 2017; Passmore u. a., 2017). Da Menschen unterschiedliche Körpergrößen haben, ist es schwierig, solche Ratschläge umzusetzen. „The Cinematic VR Field Guide“ (Jaunt, 2017) empfiehlt eine Höhe von ca. 177 cm. In Szenen mit Protagonisten kann deren Augenhöhe eine hilfreiche Referenz sein, wie dies auch bei framed Filmen der Fall ist.

Neben dieser visuellen Quelle sensorischer Informationen ist die Posture-Information wichtig. Diese Information bezieht der Betrachter aus der Kenntnis, über die eigene Haltung (z. B. Sitzen oder Stehen), was die Wahrnehmung der CVR-Experiences beeinflussen könnte. Da Zuschauer den Film in verschiedenen Körperhaltungen ansehen, z. B. Sitzen oder Stehen und die Aufnahme in der Regel nur in einer Höhe aufgenommen wird, muss sich der Filmemacher für eine Kameraposition entscheiden.

3.2.1 Diskrepanz in der Körpergröße

Um zu untersuchen, ob die Höhe der Kamera Presence, Simulator-Sickness und Benutzererfahrung beeinflusst, führten wir drei Experimente durch.

Ziel des ersten Experimentes in **Studie 7.2.3** war es, herausfinden, ob das Gefühl des Benutzers für die eigene Körpergröße in der virtuellen Realität der tatsächlichen Körpergröße entspricht. Zu diesem Zweck wurde eine virtuelle Umgebung geschaffen, die einige charakteristische Möbel in standardisierten Größen enthielt. Der Benutzer konnte die Höhe der virtuellen Kamera ändern, bis sie dem Gefühl entsprach, in der virtuellen Welt die eigene reale Körpergröße zu haben. Die Ergebnisse wiesen eine Tendenz auf, dass Menschen ihre eigene Körpergröße unterschätzen. Dies kann durch die Tatsache verursacht sein, dass Menschen es eher gewohnt sind, die Welt aus einer niedrigeren Perspektive zu sehen, und diese aus dieser Perspektive daher vertrauter ist. Wir sitzen normalerweise für lange Zeiträume des Tages. Als Kinder waren wir es gewohnt, kleiner zu sein und eine niedrigere Augenhöhe zu haben.

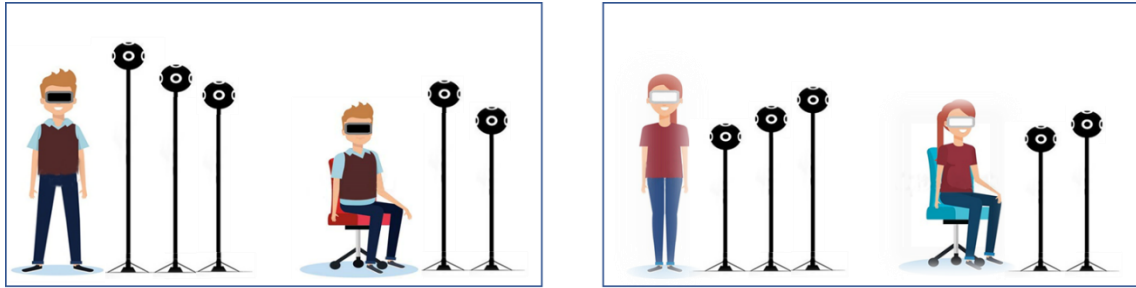


Abbildung 3-7: Testfälle für die beiden Gruppen, links: above-Gruppe, rechts: below-Gruppe.

Im zweiten Experiment der **Studie 7.2.3** wurde untersucht, ob kleinere Unterschiede (10 cm, 20 cm) zwischen physischer und virtueller Augenhöhe den Betrachter irritieren. Dabei wurden sowohl sitzende wie auch stehende Positionen betrachtet und zwei unterschiedliche Gruppen miteinander verglichen: Eine Gruppe schaute sich Videos an, deren Setups höher oder gleich ihrer eigenen Augenhöhe waren (*above*-Gruppe, Abbildung 3-7, links). Die andere Gruppe schaute Videos, deren Kamera niedriger oder gleich ihrer eigenen Größe waren (*below*-Gruppe, Abbildung 3-7, rechts).

Diskussion

Vergleicht man die Presence für die verschiedenen Kamerahöhen, deuten die Ergebnisse auf eine höhere Presence in der *below*-Gruppe hin. Dieses Ergebnis trat sowohl sitzend als auch stehend auf.

Die meisten Teilnehmer der *above*-Gruppe konnten erkennen, welches Video auf ihrer tatsächlichen Augenhöhe aufgenommen wurde, und bevorzugten dieses Video. In der sitzenden Position identifizierten sogar alle Teilnehmer das zu ihrer eigenen Körpergröße gehörende Video. Dagegen identifizierten in der *below*-Gruppe weniger Menschen das Video mit ihrer eigenen Größe. Der Unterschied nach unten scheint weniger auffallend zu sein als der nach oben.

Zusätzlich hatten in beiden Postures (Sitzen/Stehen) fast doppelt so viele Teilnehmer unwohle Gefühle, wenn sie das Video mit der zu hohen Kameraposition sahen. Es ist anzunehmen, dass eine zu niedrige Kamera einer zu hohen Kamera vorgezogen werden sollte.

Die meisten Teilnehmer, die die Videos aus der Perspektive über der eigenen Augenhöhe sahen, gaben ein unwohles Gefühl an. In dem Fall, dass die Kamera zu niedrig war, wurde dies von weniger Teilnehmern beanstandet. Ein Grund dafür könnte wiederum sein, dass wir daran gewöhnt sind, auch niedrigere Augenpositionen einzunehmen.

Wir hatten keinen Unterschied in der Baseline (Kamera auf Augenhöhe) für die beiden Gruppen erwartet. Allerdings schnitten Presence, Aufmerksamkeit und Wohlbefinden in der Baseline besser ab, wenn die Zuschauer sich zuvor ein Video angesehen hatten, welches zu hoch aufgenommen war. Möglicherweise findet eine Art „innere Kalibrierung“ statt. Eine mögliche Erklärung wäre: Da eine hohe Kamera als zu hoch bemerkt wird, wird die danach folgende Kamera auf Augenhöhe als angenehm empfunden. Dagegen wird die zu niedrige Kamera als Baseline angenommen und die darauffolgende Kamera auf Augenhöhe dann als zu hoch empfunden. Um diese Vermutung zu evaluieren bedarf es allerdings weiterer Studien.

3.2.2 Diskrepanz in Postures

Da der Filmemacher nicht weiß, ob der Zuschauer den Film sitzend oder stehend sehen wird, ist es wichtig zu wissen, wie ein in Sitzhöhe aufgenommener Film auf einen stehenden Zuschauer wirkt - und umgekehrt. Dazu sahen sich die Teilnehmer in **Studie 7.2.4** das in Standhöhe aufgenommene Video sitzend an (Abbildung 3-8, links) und das in Sitzhöhe aufgenommen Video im Stehen (Abbildung 3-8, rechts).

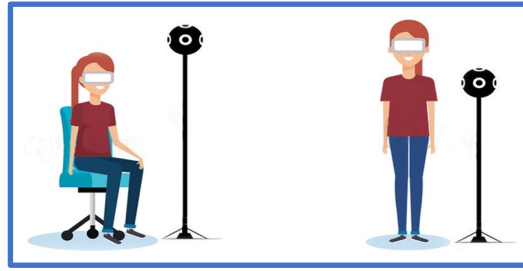


Abbildung 3-8: Sitzend schauen die Teilnehmer einen Film an, der in Stehhöhe aufgenommen wurde, und stehend einen Film in Sitzhöhe.

Für die Teilnehmer war es einfacher, eine Sitzhaltung an eine virtuelle Stehposition anzupassen als umgekehrt. Dies steht im Einklang mit anderen Forschungen, die den Einfluss der Augenhöhe in virtuellen Umgebungen untersuchten (Leyrer u. a., 2011). Unsere Ergebnisse stimmen mit der Aussage überein, dass eine Zunahme der Augenhöhe zu einer Abnahme der wahrgenommenen Entfernung führt, eine Abnahme der Augenhöhe jedoch keinen Einfluss auf die wahrgenommene Entfernung hat (Leyrer u. a., 2011). Eine zu geringe Augenhöhe scheint ein geringeres Problem zu sein als eine zu hohe Augenhöhe. Dafür gibt es mehrere Erklärungen. Einerseits verändern Menschen im täglichen Leben sehr oft ihre Augenhöhe und sind es gewohnt, ihre Umgebung aus niedrigeren Positionen zu beobachten. Es ist jedoch selten, die Umgebung von einer Position aus zu erleben, die höher als die eigene Körpergröße ist.

In unserer Studie hatten die Betrachter den Wunsch, ihre Posture zu ändern und sich zu setzen, wenn sich die Kamera auf Sitzlevel befand. Weniger Teilnehmer wollten aufstehen, als sie ein Video in Stehhöhe anschauten. Die Tatsache, dass das Sitzen angenehmer ist als das Stehen, könnte das Ergebnis beeinflusst haben. Darüber hinaus sind die Zuschauer es gewohnt, Filme im Sitzen anzuschauen. Die Videos waren sehr kurz, so dass diese Überlegung keinen großen Einfluss haben sollte, aber wahrscheinlich würden noch mehr Leute es vorziehen, sich hinzusetzen, wenn die Videos länger sind. Der Trend, sich auch für ein kurzes Video hinzusetzen, könnte jedoch für alle Höhen von Kameras gelten. Bei weiteren Untersuchungen sollte dies berücksichtigt werden und mehr Wert auf sitzende Zuschauerpositionen gelegt werden.

Die Ergebnisse zeigen, dass der Betrachter in der Lage ist, eine Sitzhaltung zu adaptieren und eine Szene zu akzeptieren, selbst wenn diese in Standhöhe aufgenommen wurde. Schwieriger dagegen ist es, in einer Standposture eine Sitzhöhe zu assimilieren. Daher ist es vorzuziehen, für allgemeine Fälle Standhöhen zu verwenden. Diese Position ist für die meisten Benutzer sowohl im Stehen als auch im Sitzen angenehm. Es ist also nicht notwendig, sich nach der sitzenden Position des Zuschauers zu richten. Erfordert allerdings die Geschichte eine Sitzhöhe (z. B. sitzen an einem Tisch), sollte diese verwendet werden.

3.2.3 Fazit

Zusammenfassend wurden die folgenden Erkenntnisse gefunden:

- Unterschiede zwischen Kamera- und Augenhöhe werden eher akzeptiert, wenn die Kameraposition niedriger ist als die eigene Augenhöhe.
- Sitzpositionen werden bevorzugt und können einfacher an die wirkliche Posture angepasst werden als Stehpositionen.
- Benutzer bevorzugen es, sich beim Anschauen von Videos hinzusetzen, insbesondere, wenn die Kamerahöhe der Augenhöhe einer sitzenden Person entspricht.
- In Stehhöhe aufgenommene Videos sind für die meisten Benutzer sowohl in Sitzpositionen als auch im Stehen angenehm. Es scheint jedoch schwieriger zu sein, ein Video, welches in Sitzhöhe aufgenommen wurde, stehend anzusehen.

Wie unsere Ergebnisse zeigen, können die meisten Zuschauer ihre Sitzhaltung an eine Kamerahöhe im Stehen anpassen. Dies entspricht dem Ergebnis von Leyrer et al. (Leyrer u. a., 2015) über die Bedeutung von Posture-Informationen für die Bestimmung der Augenhöhe. Unter der Annahme, dass die Zuschauer beim Anschauen von Filmen sitzen, können die Filmemacher die Kameras je nach Story sowohl in Sitzhöhe als auch in Stehhöhe platzieren. Diese Ergebnisse können zur Verbesserung der Guidelines für CVR-Kamerapositionen verwendet werden.

3.3 Kamerapositionen in CVR

3.3.1 Kameraabstände und Einstellungsgrößen in CVR

Auch wenn in CVR der Filmemacher die Kadrierung nicht genau festlegen kann, kann die Einstellungsgröße durch den Kameraabstand beeinflusst werden. Um Richtlinien für geeignete Abstände zu finden, wurden die Erkenntnisse der Proxemik verwendet. Die Abstände der Proxemik haben in der Realität eine ähnliche Bedeutung wie die Einstellungsgrößen beim Filmemachen: extreme Nahaufnahmen kennzeichnen intime Situationen, Nahaufnahmen zeigen Emotionen einer Person, und Totalen eine Person in größerer Distanz. Betrachtet man die Beschreibungen der verschiedenen Proxemik-Räume und Einstellungsgrößen, erkennt man ähnliche Bedeutungen. Es stellt sich die Frage, ob die Abstände der Proxemik-Räume Kameraabständen und Einstellungsgrößen in Cinematic VR entsprechen.

Dazu bestimmten wir als Erstes das wirklich sichtbare FoV des HMDs. Für unsere Experimente wurde die Samsung Gear VR verwendet, da mit diesem Gerät CVR-Filme im privaten Bereich angeschaut werden. Fünf Teilnehmer untersuchten drei verschiedene CVR-Szenen, um herauszufinden, wie oft ihr FoV in vertikaler und horizontaler Richtung in den sphärischen Film passt. Da die Grenzen eines FoV nicht klar erkennbar sind, kann dies nur näherungsweise erfolgen. Die Ergebnisse für alle Teilnehmer waren jedoch ähnlich und entsprachen nicht den theoretischen Angaben des FoV. Das FoV in horizontaler Richtung passte 5.5-mal in die 360°-Ansicht, in vertikaler Richtung 3-mal in den 180°-Bereich, was ein subjektives FoV von 65.5° in horizontaler und 60° in vertikaler Richtung bedeutet (Abbildung 3-9).

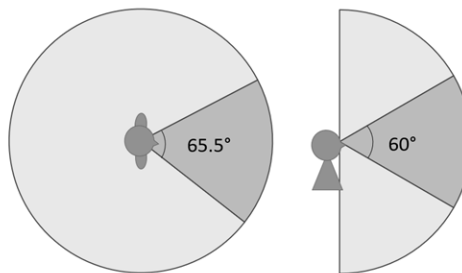


Abbildung 3-9: Das experimentell ermittelte FoV in horizontaler (links) und vertikaler (rechts) Richtung.

Diese Ergebnisse hängen in hohem Maße vom verwendeten HMD ab. Die technische Entwicklung bringt neue Geräte mit breiteren FoVs auf den Markt. Möglicherweise werden HMDs in naher Zukunft das augenbasierte Sichtfeld nicht einschränken. Trotzdem sind die berechneten Werte ideal für unsere Forschung geeignet, da sie dem für die Bilderkennung verantwortlichen Bereich des FoV ähneln, der ungefähr 60° beträgt (siehe 3.1.3).

Mit dem ermitteltem FoV lassen sich die maximalen Objektgrößen berechnen, die im FoV sichtbar sind, wenn der Betrachter (die Kamera) in verschiedenen Abständen positioniert ist.

$$\text{Breite} = 2d \cdot \tan(\alpha/2) = 2d \cdot \tan(65.5^\circ/2) = 2d \cdot 0.64 = 1.28d$$

$$\text{Höhe} = 2d \cdot \tan(\beta/2) = 2d \cdot \tan(60^\circ/2) = 2d \cdot 0.575 = 1.15d$$

Dabei ist d die Kameraentfernung, α das horizontale und β das vertikale FoV.

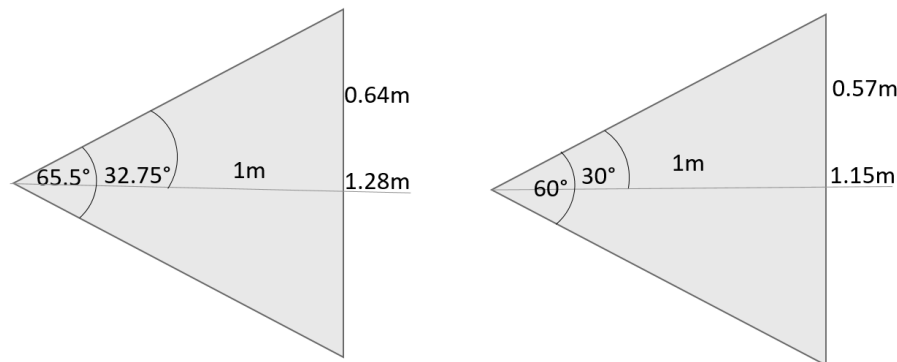


Abbildung 3-10: Die Abbildung zeigt die Breite (links) und Höhe (rechts) eines Objekts, das im FoV sichtbar ist, wenn der Betrachter (Kamera) 1 m entfernt ist.

In Abbildung 3-10 werden die Werte für eine Kameraentfernung von 1 m illustriert. Die Werte für die Grenzen der Proxemik-Zonen werden in Tabelle 3-1 aufgeführt.

	Bereich	Breite im FoV	Höhe im FoV	Einstellung
intime Zone	0.45	0.58	0.52	nah
persönliche Zone	1.2	1.54	1.38	halbnah
soziale Zone	3.6	4.61	4.14	halbtotale
öffentliche Zone	7.6	9.73	8.74	total

Tabelle 3-1: Größen der Proxemik-Spaces (in Meter) und maximale Ausdehnung der im FoV sichtbaren Bereiche von Objekten, die in den Spaces platziert sind.

Die berechneten Größen wurden verwendet, um die Einstellungsgrößen den Proxemik-Räumen zuzuordnen. Dabei wurde von einer ungefähren Körpergröße von 1.70 m ausgegangen. Da sowohl Proxemik-Abstände, Körpergrößen wie auch die Grenzen der Einstellungsgrößen Werte sind, die in bestimmten Bereichen variieren, beschreiben die berechneten Werte ebenfalls keine exakten Grenzen und werden grafisch in Abbildung 3-11 dargestellt.

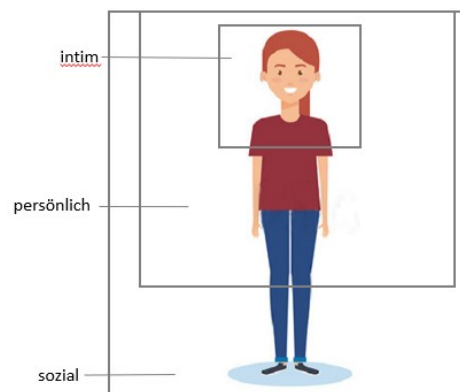


Abbildung 3-11: Einstellungsgrößen, die den verschiedenen Proxemik-Räumen zugeordnet sind.

Die Ergebnisse der **Studie 7.2.1** bestätigen, dass Proxemik-Distanzen ähnliche Wirkungen beim Zuschauer hervorrufen wie Einstellungsgrößen. In einer Museums Umgebung, in der eine Protagonistin weitergehende Informationen vermittelte, empfanden die Zuschauer die *persönliche* Distanz (1 m) am

natürlichsten und angenehmsten. Bei den beiden nahen Distanzen (*intim, persönlich*) hatten die Teilnehmer ein intensiveres Gefühl, dass die Person im Video sie ansieht und zu ihnen spricht, als in den beiden entfernten Distanzen (*sozial, öffentlich*). Die *intime* Distanz wirkte für das gewählte Szenarium zu distanzlos.

3.3.2 F-Formationen in CVR

Neben den Entfernungen ist es wichtig, wie Kamera und Charaktere zueinander angeordnet sind. In CVR unterscheidet sich die Aufteilung des Raumes vom konventionellen Filmemachen. Während für einen framed Film die Filmszene in einem begrenzten Bereich arrangieren wird, müssen CVR-Filmemacher über den gesamten Raum nachdenken. Außerdem entscheidet der konventionelle Filmemacher, ob zwei Personen gemeinsam in einem Bild oder getrennt in zwei aufeinanderfolgenden Bildern zu sehen sind. In CVR können die Benutzer diese Tatsache durch ihre eigene Blickrichtung beeinflussen. Der Filmemacher kann jedoch entscheiden, ob Personen so nah beieinander sind, dass sie gleichzeitig wahrgenommen werden können, oder so weit voneinander entfernt sind, dass der Betrachter den Kopf bewegen muss, um beide zu sehen.

In **Studie 7.2.2** wurde eine Gesprächssituation zwischen zwei Personen untersucht und dabei verschiedene F-Formationen von zwei Charakteren und einer Kamera untersucht (Abbildung 3-12): die *joint-Seitenansicht* (beide Personen gleichzeitig von der Seite sichtbar), die *split-Seitenansicht* (Personen nur getrennt sichtbar, von der Seite), die *Frontalansicht* (Zuschauer steht genau zwischen Charakteren).

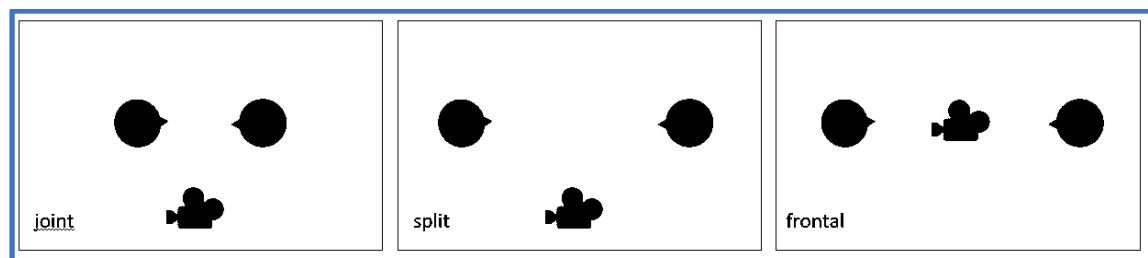


Abbildung 3-12: Testfälle von links nach rechts: joint: beide Personen gleichzeitig von der Seite sichtbar, split: Personen nur getrennt sichtbar (von der Seite), frontal: Zuschauer steht genau zwischen Charakteren.

Die Ergebnisse zeigen, dass die Teilnehmer die wichtigsten Informationen (die beiden Charaktere) lieber so nah beieinander haben, dass sie gleichzeitig im FoV sichtbar sind. Einige Teilnehmer bevorzugten jedoch auch die unrealistische Situation zwischen den Charakteren.

Die Ergebnisse der Benutzerstudien resultieren in den folgenden Erkenntnissen:

- Proxemik-Distanzen können auf Abstände zwischen Kamera und Charaktere übertragen werden.
- Wenn die Charaktere nicht gleichzeitig gesehen werden können, wechseln die Zuschauer häufig die Ansicht zwischen sprechenden Personen.
- Es ist zu beachten, dass der Betrachter ein Gesicht oft nicht mittig im FOV zentriert. Stattdessen wird oft intuitiv ein Bildausschnitt, der den Regeln des Goldenen Schnittes folgt, ausgewählt.

Die Ergebnisse unserer Studie repräsentieren das aktuelle Verhalten der Zuschauer, das sich mit zunehmenden VR-Erfahrungen ändern kann. Es ist jedoch wichtig zu wissen, wie sich die Zuschauer in verschiedenen Szenarien fühlen, um eine Filmsprache für CVR zu entwickeln.

3.3.3 Fazit

In diesem Abschnitt wurde eine Einordnung der Kameraentfernung und Einstellungsgrößen basierend auf den Kenntnissen der Proxemik vorgestellt. Es wurden die Abstände der Proxemik in Kameraentfernungen

übertragen und gezeigt, dass sie den in framed Filmen benutzten Einstellungsgrößen entsprechen. Neben dem Abstand der Kamera zu den Charakteren sind auch Positionen und Abstände der Charaktere zueinander von Bedeutung. In einem Experiment konnten erste Erfahrungen gesammelt werden, um F-Formationen in CVR zu nutzen. Da das Gebiet der F-Formationen sehr komplex ist und viele Konstellationen existieren, sind weitere Experimente notwendig, um allgemeingültige Aussagen machen zu können.

3.4 Kamerabewegungen in CVR

Das Schwenken der Kamera ist in CVR nicht erforderlich, da dem Zuschauer das gesamte sphärische Bild zur Verfügung steht und Schwenks von ihm selbst ausgelöst werden. Eine Drehung der Kameraansicht, die von außen initiiert wird, kann zu Simulator-Sickness führen, da die tatsächliche Kopfbewegung des Betrachters nicht mit der Drehung in der virtuellen Welt korrespondiert (Davis u. a., 2014).

Im Gegensatz zur Drehung der virtuellen Kamera, die der CVR-Betrachter selbst ausführen kann, ist es in 3DoF-Umgebungen nicht möglich den Platz durch eigene Fortbewegung selbst zu ändern. Es ist deshalb wichtig zu untersuchen, wie eine bewegte Kamera auf einen sich nicht bewegenden Zuschauer wirkt.

3.4.1 Kamerafahrten in CVR

In **Studie 7.2.5** wurde eine Kamerabewegung (*move*) mit einer Sequenz von geschnittenen Aufnahmen (*cut*) und einer Kombination aus beidem (*mix*) verglichen (Abbildung 3-13). In allen drei Testfällen wurde das gleiche Material mit der gleichen Story verwendet: Es gibt fünf unterschiedliche Aktionsorte, die entweder mit Kamerafahrt oder durch harte Schnitte miteinander verbunden sind. Dabei wurde nur die Position der Kamera geändert. Die Blickrichtung konnte vom Benutzer selbst durch Kopfbewegungen bestimmt werden.

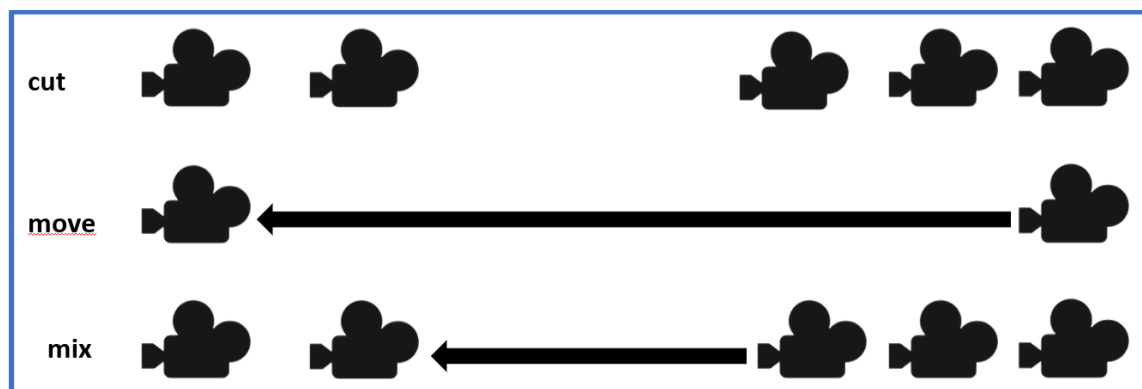


Abbildung 3-13: Darstellung der untersuchten Techniken; oben (*cut*): Fünf Kamerapositionen werden nacheinander angezeigt. Mitte (*move*): Die Kamera bewegt sich langsam von einer Kameraposition zur nächsten. unten (*mix*): Zwischen der dritten und vierten Kameraposition bewegt sich die Kamera, alle anderen Positionswechsel sind ohne Übergänge.

Die Teilnehmer der Studie bevorzugten die Kamerabewegung (*move*). Sie konnten sich besser orientieren, waren aufmerksamer und empfanden die virtuelle Welt als realistischer. Dagegen führte die gemischte Methode, in der innerhalb der Szene zwischen Kamerabewegung und Schnitt gewechselt wurde zu Orientierungslosigkeit und Schwierigkeiten bei Konzentration und Fokussierung.

Dass die *move*-Methode besser als die *cut*-Methode abschnitt, könnte durch die Orientierung beeinflusst worden sein. Da diese von uns nicht geändert wurde, musste der Betrachter nach dem Schnitt erst selbst die Rol finden. Auch wenn die Orientierung der vor dem Schnitt entsprach, scheint dies im Fall von

timeline-basierten Schnitten (Schnitte zu einem bestimmten Timecode) die Zuschauer zu verwirren, wie auch in **Studie 7.3.1** festgestellt wurde. In **Studie 7.3.6** wurde gezeigt, dass *toRoI*-Schnitte in diesen Fällen besser funktionieren.

3.4.2 Fazit

Auch wenn eine Kamerabewegung (virtuell) nicht der realen Bewegung eines CVR-Betrachters entspricht, werden solche Bewegungen vom Betrachter akzeptiert, wenn sie reibungslos und langsam sind. Sie wurden in verschiedenen CVR-Applikationen bereits realisiert (Crew United, 2019; Oculus, 2018).

3.5 Guidelines für Kameraführung in CVR

Aus den gewonnenen Erkenntnissen leiten wir im Folgenden Guidelines zur Kameraführung in CVR her.

Abstände

Distanzen aus der Proxemik sind ein Äquivalent zu Einstellungsgrößen in framed Filmen und können in Abständen zwischen Kamera und Charakter übertragen werden. Für die Umsetzung von Einstellungsgrößen können die Distanzen in Tabelle 3-1 als Kameraabstände genutzt werden.

Zuschauer bevorzugen zwei miteinander sprechende Charaktere so nah beieinander zu haben, dass diese gleichzeitig im FoV sichtbar sein können. Einige Teilnehmer zogen jedoch die unrealistische Situation zwischen den Charakteren vor, da sie sich so in der virtuellen Welt präsenter fühlten.

Höhen

In Stehhöhe aufgenommene Videos sind für die meisten Benutzer sowohl in Sitzpositionen als auch im Stehen angenehm. Zuschauer können ihre momentane Körperhaltung (Sitzen/Stehen) adaptieren. Es scheint jedoch angenehmer zu sein, ein Video, aufgenommen auf Stehhöhe, sitzend anzusehen als ein Video in Sitzhöhe stehend. Unter der Annahme, dass Zuschauer beim Anschauen von Filmen meist sitzen, können die Filmemacher die Kameras je nach Story sowohl auf Sitz- wie auch Stehhöhe platzieren.

Bei Unterschieden zwischen Kamera und Augenhöhe, die nicht durch die Körperhaltung verursacht sind, haben wir festgestellt, dass Kamerapositionen, die unter der Augenhöhe des Betrachters liegen, weniger Schwierigkeiten bereiten als Kamerapositionen über der Augenhöhe. Dementsprechend sollten Filmemacher die Kamera eher zu tief als zu hoch platzieren. Unter Berücksichtigung unterschiedlicher Körpergrößen sollten die Kamerapositionen auf kleinere Personen ausgerichtet sein, da diese auch von größeren Personen leichter adaptiert werden können.

Basierend auf einer durchschnittlichen Körpergröße von 1,76 m (Männer) und 1,62 m (Frauen)(National Center for Health Statistics, 2017) und einer durchschnittlichen Augenhöhe von 1,66 m (Männer) und 1,52 m (Frauen) ist eine Kamerahöhe von 1,56 m für stehende Situationen empfehlenswert.

Kamerabewegungen

Kamerafahrten werden vom Betrachter akzeptiert, wenn sie gleichmäßig und langsam sind. Schwenks werden vom Zuschauer selbst durch Kopfbewegungen ausgeführt.

Zusammenfassung

Zusammenfassend wurden die folgenden Guidelines gefunden, die allerdings nicht als starre Regeln zu verstehen sind. Sie basieren auf Experimenten mit ausgewählten Filmsequenzen, die nicht die Fülle der möglichen Filmstile abdecken. Auch bewusste Verstöße gegen diese Hinweise, können in CVR-Experiences eingebaut werden, um spezifische Emotionen hervorzurufen.

Abstände: Distanzen aus der Proxemik können in Abstände zwischen Kamera und Charakter übertragen werden (Distanzen: Tabelle 3-1).

Höhen: Die Kamera kann je nach Story sowohl auf Sitz- wie auch Stehhöhe platziert werden und sollten sich dabei an den Augenhöhen kleinerer Personen orientieren.

Kamerafahrten werden vom Betrachter akzeptiert, wenn sie gleichmäßig und langsam sind.

Schwenks werden vom Zuschauer selbst durch Kopfbewegungen ausgeführt.

Die Ergebnisse unserer Studie repräsentieren das aktuelle Verhalten der Zuschauer, welches sich mit zunehmenden VR-Erfahrungen ändern kann. Es ist wichtig zu wissen, wie sich die Zuschauer in verschiedenen Szenarien fühlen, um eine Filmsprache für CVR zu entwickeln. Dieser langfristige Prozess ist nicht abgeschlossen und wird sich immer wieder an neue Techniken und Sehgewohnheiten anpassen.

Diese Guidelines sind nur Empfehlungen und ein erster Schritt, um die Wirkung verschiedener Kamerapositionen zu nutzen. Neue technische Verfahren und die Entwicklung unseres Sehverhalten werden weitere Möglichkeiten hervorbringen.

3.6 Fazit und Ausblick

In diesem Kapitel wurde Kamerapositionen und -bewegungen untersucht. Es wurde festgestellt, dass die meisten Zuschauer in der Lage sind, Kamerahöhen, die nicht ihrer eigenen Augenhöhe entsprechen, zu akzeptieren. Niedrigere Kamerahöhen bereiten dabei wenige Probleme und sind oft nicht von der korrekten Höhe zu unterscheiden. Kamerabewegungen sind auch in CVR möglich, wenn sie langsam und vorhersehbar (geradlinig) sind. All unsere Untersuchungen sind eine erste Annäherung an dieses Thema und die Ergebnisse sind von unserem verwendeten Material beeinflusst. Weitere Forschungen sind notwendig, um mehr Erkenntnisse zu diesem Thema zu finden. Dabei sollten unter anderem längere Filme, stereoskopische Sequenzen, sich bewegende Charaktere und verschiedene Filmgenres mit tragenden Stories näher betrachtet werden.

Die Platzierung der Kamera ist nur ein Instrument, das zusammen mit allen anderen Stilmitteln bei der Entwicklung einer Filmsprache für CVR berücksichtigt werden muss. Unsere Forschung ist ein Schritt in diese Richtung.

4

How to Cut: Montage in CVR mittels Spaceline-Konzept

„Die erste besteht darin, daß das Bündel dieser Aufsätze auf gar keinen Fall nacheinander betrachtet und rezipiert werden soll. Ich wünschte mir, daß sie alle zugleich wahrgenommen werden könnten, weil sie schließlich eine Reihe von Sektoren darstellen, die, auf verschiedene Gebiete ausgerichtet, um einen allgemeinen, sie bestimmenden Standpunkt - die Methode - angeordnet sind. Andererseits wollte ich rein räumlich die Möglichkeit schaffen, daß jeder Beitrag unmittelbar mit einem anderen in Beziehung tritt, daß einer in den anderen übergeht. Daß sie sich wechselseitig aufeinander berufen. Einer den anderen ergänzt. Solcher Synchronität und gegenseitigen Durchdringung der Aufsätze könnte ein Buch in Form ... einer Kugel Rechnung tragen! Wo alle Sektoren der Kugel auf einmal präsent sind, und egal, wie weit sie voneinander entfernt sind – immer ist ein direkter Übergang von einem zum anderen über das Zentrum der Kugel möglich. Aber leider ... werden Bücher nicht als Kugeln geschrieben.“

(Sergej Eisenstein, Tagebuch 5. August 1929)

Die Montage ist ein wichtiges stilistisches Mittel und prägt einen Film wesentlich. Mittels Schnitten ermöglicht sie, zwischen Orten und Charakteren zu wechseln, zeitliche Abläufe zu komprimieren, Kameraperspektiven und Einstellungsgrößen zu ändern. Da sich in CVR der Zuschauer selbst in der Szene umschaute, benötigt der Einsatz von Schnitten neue Herangehensweisen. Einerseits verlagert sich ein wichtiger Teil der Montage – die Auswahl des sichtbaren Bildausschnittes – vom Filmmacher zum Zuschauer, andererseits eröffnen sich durch die zusätzlich zur Verfügung stehende Raumkomponente neue Möglichkeiten. Die Blickrichtung des Zuschauers kann genutzt werden, um neue Filmkonstruktionen umzusetzen. In dem nachfolgend dargelegten Spaceline-Konzept, basiert ein Schnitt nicht mehr nur auf der verstrichenen Zeit, sondern auch auf der Blickrichtung des Zuschauers. In Anlehnung an den Begriff Timeline wird der Begriff Spaceline eingeführt. Während auf einer Timeline Filme in ihrem zeitlichen Ablauf montiert werden, beschreibt ein Spaceline-Konstrukt die räumlichen Zusammenhänge. Dazu braucht es eine Spezifikation der Regionen und Szenenübergangsereignisse ebenso wie Techniken für die Transitions.

Dieses Kapitel enthält bereits veröffentlichtes Material aus folgenden Publikationen:

- Rothe, S., Hussmann, H., 2019. Spaceline: A Concept for Interaction in Cinematic Virtual Reality. in: Interactive Storytelling - 11th International Conference on Interactive Digital Storytelling - ICIDS 2019.
- Rothe, S., Brunner, H., Buschek, D., Hussmann, H., 2018. Spaceline: A Way of Interaction in Cinematic Virtual Reality. in: Symposium on Spatial User Interaction - SUI'18.
- Rothe, S., Brunner, H., Hussmann, H., 2018. Spaceline und Interaktion in Cinematischer Virtueller Realität. In: Mensch und Computer - MuC '18
- Rothe, S., Pothmann, P., Drewes, H., & Hussmann, H. Interaction techniques for cinematic virtual reality. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2019.

Ein Prototyp des CVR-Editors wurde von Harald Brunner im Rahmen seiner Masterarbeit implementiert und für die Durchführung der Studien 7.3.3 und 7.3.4 verwendet.

In diesem Kapitel wird das neue Konzept einer Spaceline vorgestellt. Es berücksichtigt zur Bestimmung von Anfang und Ende eines Filmsegmentes nicht nur die Zeit – wie in Timelines – sondern die Blickrichtung des Zuschauers, wodurch sich neue Möglichkeiten der Blicklenkung und Interaktion ergeben.

Dafür werden in Abschnitt 4.1 Erkenntnisse aus den Gebieten Filmtheorie, Psychologie und Mensch-Computer-Interaktion dargestellt und diskutiert.

Für die Präsentation des Spaceline-Konzeptes in Abschnitt 4.2 werden grundlegende Begriffe eingeführt und die Unterschiede zwischen den Terminologien von Timeline- und Spaceline-Konzept erläutert. Die notwendigen Methoden für die Selektion von Spaces werden vorgestellt, miteinander verglichen und ihre Vor- und Nachteile herausgearbeitet (Abschnitt 4.2.2). Um Applikationen mit dem Spaceline-Konzept umzusetzen, braucht es neue Montagetechniken, die die zusätzliche Raumkomponente berücksichtigen und die Szenen nach unterschiedlichen Kriterien ausrichten (Abschnitt 4.2.3).

Um praktische Unterstützung bei der Verwendung des Spaceline-Konzeptes zu geben, werden in Abschnitt 4.3 Guidelines für Spaceline-Konstrukte dargelegt (4.3.1), die Erkenntnisse in einem Design-Space für Spaceline-Konstrukte zusammengefasst (4.3.2) und ein Konzept für ein Editor-Tool vorgestellt (4.3.3).

4.1 Theoretischer Hintergrund

In den Anfängen der Filmgeschichte bestanden Filme zunächst aus nur einer Einstellung. Es gab keine Änderungen von Perspektiven, Kamerawinkeln und Aufnahmegrößen, genauso, wie es die Zuschauer vom Theater her gewohnt waren. Allerdings bestehen auch diese ersten Filme aus vielen unsichtbaren Schnitten, die ein Bild mit dem nächsten verbinden und so ein Filmsegment ergeben (Steinman u. a., 2000). Dieser Effekt, bei dem aufeinanderfolgende Einzelbilder zu einem Bewegungseindruck verschmelzen, wird in der Wahrnehmungspsychologie Phi-Phänomen genannt (Guilford und Helson, 1929; Steinman u. a., 2000; Wertheimer, 1938) und ist der Prototyp der Montage.

Bald wurde erkannt, dass sich die einzelnen Filmsegmente miteinander verknüpfen lassen und dabei neue Wirkungen erzielen (Thompson und Bordwell, 2010). Erste systematische Film-Experimente zur Montage wurden durchgeführt. Der Filmtheoretiker Lev Kuleshov untersuchte die Verwendung von Schnitten zur emotionalen Beeinflussung der Zuschauer und kombinierte dazu Einstellungen auf verschiedene Weise. In seinen berühmten Kuleshov-Experimenten wurden zwischen die Aufnahmen eines regungslos schauenden Schauspielers unterschiedliche emotionsauslösende Aufnahmen montiert, um den Zuschauern verschiedene Situationen zu suggerieren (Beller, 2002; Wulff, 2012d). Andere Experimente von Kuleshov verbanden durch die Montage Orte miteinander, die in der Realität weit voneinander entfernt lagen, im Film aber wie ein gemeinsamer Ort wirkten (Beller, 2002). Dieser Effekt wird auch heute

noch oft verwendet und in der Filmtheorie als *kreative Geographie* bezeichnet (Amann, 2012). Damit solche Effekte funktionieren, ist es notwendig, sie in eine Geschichte einzubetten (Cutting, 2007). Erst der Kontext ermöglicht es dem Zuschauer, die Szenen auf die gewünschte Weise zu interpretieren.

Schritt für Schritt wurde eine Filmsprache entwickelt und mit Schnitten Veränderungen in Raum und Zeit umgesetzt. Filmemacher und Zuschauer lernten damit umzugehen, und die Verwendung von Schnitten wurde zu einem der wichtigsten Stilmittel: Der Schnitt bestimmt den Filmrhythmus, lenkt den Betrachter und ist Teil des Erzählens. Schnitte können Diskontinuitäten in Raum, Zeit oder Aktion überbrücken (Smith und Henderson, 2008; Smith und Martin-Portugues Santacreu, 2017) und werden benötigt, um verschiedene Einstellungsgrößen wie Totale, Halbtotale und Nahaufnahmen umzusetzen.

In CVR sehen die Zuschauer einen sphärischen Film über HMDs und fühlen sich in die Szenerie versetzt. Trotzdem können Schnitte eingesetzt werden (Kjær u. a., 2017; Moghadam und Ragan, 2017) und transportieren den Zuschauer meist an einen anderen Ort. Die Verwendung herkömmlicher Filmbearbeitungsmethoden und -werkzeuge führt zu zwei Problemen. Einerseits ist es möglich, dass Dinge, die für die Geschichte wichtig sind, vom Betrachter nicht wahrgenommen werden, da sie sich nicht in seinem FoV befinden. Andererseits wechselt die Szene nach einer gewissen Zeit, auch wenn der Betrachter sich weiter umschauchen möchte. Die Interessen und Bedürfnisse der Zuschauer sind unterschiedlich und es sollte die Möglichkeit bestehen, dass diese berücksichtigt werden.

Der Begriff *Schnitt* wird oft in unterschiedlichen Bedeutungen verwendet: zum einen wird damit der Übergang zwischen zwei Filmsegmenten bezeichnet (Hüning, 2012a), zum anderen die Phase der Filmproduktion, in der aus dem gedrehten Material ein Film zusammengefügt wird (Hüning, 2012b). Das Wort Schnitt stammt noch aus einer Zeit, in der die Aufnahmen tatsächlich an der passendsten Stelle zerschnitten wurden und geben heute nicht mehr die eigentliche Funktion wieder. Wir benutzen deshalb in CVR den Begriff *Link* - der allgemein für die Verknüpfung digitaler Daten verwendet wird - für die Verbindung zwischen zwei Filmsegmenten. Um die Unterbrechung von Filmsegmenten zu beschreiben, im Kontext von framed Filmen und bei der Analyse von Methoden aus der Literatur verwenden wir weiterhin den Begriff Schnitt (cut), um mit den in den Veröffentlichungen verwendeten Begriffen zu korrespondieren.

In der traditionellen Filmproduktion werden unterschiedliche Begriffe für die Phase des Zusammensetzens der Filmsegmente verwendet – *Schnitt* und *Montage* – wobei keine einheitliche klare Trennlinie existiert. Oft wird mit *Schnitt* der handwerkliche und mit *Montage* der kreative, künstlerische Prozess bezeichnet (Wulff, 2014). Aus den angeführten Gründen und da diese beiden Prozesse schwer trennbar sind, werden wir die Phase, die sich mit der Verbindung von Filmsegmenten auseinandersetzt, in CVR *Montage* nennen.

Für die Montage sind spezielle Kenntnisse über die Wirkung von Aufnahmen erforderlich, die am gleichen Ort aufgenommen wurden, z. B. im gleichen Zimmer. Wir nennen solche Aufnahmen *co-located* und die Verbindung zwischen solchen Aufnahmen *co-located Link*. Im Gegensatz dazu nennen wir Aufnahmen an verschiedenen Orten *dis-located*, die Verbindung zwischen ihnen *dis-located Link*.

Derzeit verbinden die meisten Links in CVR *dis-located* Aufnahmen. Da der Betrachter den sichtbaren Bildausschnitt selbst ändert, sind häufig keine zusätzlichen Links in einer Szene erforderlich. Es gibt jedoch mehrere Gründe, für eine Segmentierung:

- Lenkung der Aufmerksamkeit auf eine Rolle
- Verwendung verschiedener Einstellungsgrößen
- Einsatz verschiedener Kameraperspektiven

Außerdem sind Links, genau wie Schnitte in framed Filmen, für den Stil eines Films relevant und beeinflussen das Seherlebnis. Dabei ist es einfacher, zwei *dis-located* Aufnahmen zu verknüpfen, da der Betrachter eine neue Umgebung sieht und es keine Konflikte in der Orientierung gibt. Wenn jedoch zwei *co-located* Aufnahmen aufeinanderfolgen, kann der Betrachter verwirrt sein, wenn sich Richtung oder Position ändern. Es ist wichtig zu wissen, welche Kamerapositionen und Richtungen für *co-located* Links verwendet werden können. Unsere Forschung ist ein erster Schritt, um dieses Problem zu untersuchen.

4.1.1 Psychologische Aspekte

Menschen teilen dynamische Ereignisse in kleinere unterscheidbare Abschnitte (Zacks u. a., 2007). Die *Event Segmentation Theory* basiert auf dem Wissen über diese Prozesse. Die Segmentierung ist wichtig für die Wahrnehmung, das Merken und die Erkennung von Ereignissen (Zacks, 2010).

Die Event Segmentation Theory kann auch auf das Schauen von Filmen angewendet werden (Cutting, 2014; Magliano und Zacks, 2011; Smith, 2006; Smith und Henderson, 2008). Smith (Smith, 2006) analysierte, wie sich die Erkenntnisse der Event Segmentation Theory nutzen lassen, um Schnitte für den Zuschauer möglichst unsichtbar zu machen. Dafür wurden vier verschiedene Positionen der Schnitte untersucht: während eines Ereignisses, gegen Ende eines Ereignisses, zwischen zwei Ereignissen, am Beginn eines neuen Ereignisses. Ein Schnitt zu Beginn eines Ereignisses stellt sicher, dass das neue und alte Ereignis vollständig verarbeitet wird, während die visuelle Störung des Bewegungsschnitts ausgeblendet wird. Dieser Effekt, der Schnitte unsichtbar macht, wird auch *Edit-Blindheit* genannt.

Magliano und Zacks (Magliano und Zacks, 2011) untersuchten den Einfluss von Kontinuität im Schnitt (continuity editing) auf die Ereignissegmentierung und identifizierten drei Kontinuitätsgrade: (1) Kontinuität in Raum, Zeit und Aktion, (2) Kontinuität in Aktion, aber Diskontinuität in Raum/Zeit, (3) Diskontinuität in Raum, Zeit und Aktion. Die Diskontinuität in Aktion hatte den stärksten Einfluss auf die Ereignissegmentierung, die Diskontinuität in Raum oder Zeit wirkte sich weniger aus.

Untersuchung dieser drei Arten von Diskontinuitäten wurden von Serrano et al. (Serrano u. a., 2017) in VR fortgesetzt. Die Ergebnisse bestätigten, dass auch hier eine Diskontinuität in der Aktion den größten Einfluss auf die Ereignissegmentierung hat. Außerdem wurden Schnitte in CVR-Videos auf die Häufigkeit der verschiedenen Diskontinuitäten untersucht und festgestellt, dass es kaum Schnitte mit Kontinuität in Raum, Zeit und Aktion (1) gibt. In den von Magliano et al. untersuchten framed Filmen, war diese Schnittart die häufigste (Magliano und Zacks, 2011). Dies liegt daran, dass diese Art von Schnitten in CVR schwieriger umzusetzen sind und auch weniger gebraucht werden, da der Zuschauer selbst den Bildausschnitt ändert, solange keine Unterbrechung der Kontinuitäten stattfindet. Alle betrachteten Schnitte waren timeline-basiert. Verbindungen von Aufnahmen mit Kontinuität in Raum, Zeit und Aktion können trotzdem auch in CVR Verwendung finden, wenn *spaceline-basierte* Transitions verwendet werden, was wir in diesem Kapitel zeigen werden.

4.1.2 Schnitt via Timeline

Schnitt in traditionellen Filmproduktionen

Schnitte haben im Film unterschiedliche Funktionen. Sie verkürzen die Zeit einer aufgezeichneten Abfolge von Aktionen, machen Ortswechsel möglich und erzeugen Spannung. Die Länge einer Einstellung beeinflusst das Tempo und den Rhythmus des Films (Bordwell und Thompson, 2013). Walter Murch (Murch, 1992) beschreibt sechs Kriterien für einen guten Schnitt und gibt in Prozent an, wie wichtig diese jeweils sind:

1. Gefühl (Emotion) – 51 %,
2. Handlung (Story) – 23 %,
3. Rhythmus – 10 %,
4. Augenspur (Eye Trace) – 7 %,
5. Zweidimensionale Fläche der Leinwand – 5 %,
6. Dreidimensionaler Handlungsraum – 3 %.

Die Prozentzahlen sind dabei nicht als exakte Werte gedacht, sondern drücken aus, dass das Gefühl (1) in einer Story wichtiger ist, als alle anderen Kriterien zusammen. Wenn ein Editor nicht alle sechs Kriterien erfüllen kann, sollte das letzte zuerst aufgegeben werden. In CVR sind das 5. und 6. Kriterium nicht trennbar, weil der gesamte dreidimensionale Raum aufgenommen wird und nicht nur ein Bildausschnitt. Das Kriterium (5) korrespondiert in CVR zu dem vom Betrachter ausgewählten Bildausschnitt und ist nicht

allein durch den CVR-Entwickler bestimmbar. Der in Kriterium (6) beschriebene Handlungsraum ist in CVR der die Kamera umgebende Raum und wird in der Regel vollständig aufgezeichnet. Da sich der Betrachter virtuell in diesem Raum befindet, sind die Kriterien (5) und (6) möglicherweise wichtiger als bei framed Filmen. Die Bedeutung der Augenspur ist abhängig von Displaygröße und Filmtempo (Toczyński, 2018). Der Rhythmus (3) kann in CVR nicht allein vom Filmemacher bestimmt werden. Er wird auch vom Betrachter durch schnelle oder langsame Kopfbewegungen beeinflusst.

Das perfekte Verbinden von Aufnahmen ist eine Kunst, die den Film erheblich beeinflusst. D'Ydewalle et al. (d'Ydewalle u. a., 1998) definierten drei Arten von Schnittfehlern:

- Fehler erster Art: kleine Verschiebungen der Kamerapositionen, d. h. kleine Änderungen der Bildinhalte (30°-Regel)
- Fehler zweiter Art: Richtungswechsel in der Rechts-Links-Orientierung (180°-Regel)
- Fehler dritter Art: Diskontinuität in Platzierung von Gegenständen oder Personen

Ein wichtiges Prinzip, die *30°-Regel*, beschreibt ein Problem erster Art (d'Ydewalle u. a., 1998; Hurbis-Cherrier, 2018): Wird das gleiche Objekt aufgenommen und befindet sich die Kameraposition nach dem Schnitt zu nah an der vorherigen (weniger als 30° Unterschied in Bezug auf das Objekt), sind die Bilder sehr ähnlich und der Schnitt verwirrt den Betrachter (Murch, 1992). Da der CVR-Betrachter die Blickrichtung frei wählt, kann eine kleine Verschiebung der Kamera zu einem Bild führen, welches sich erheblich von dem zuvor beschriebenen Bild unterscheidet, falls die Blickrichtung geändert wurde. Ohne die Blickrichtung des Betrachters zu kennen, ist es schwierig, die 30°-Regel einzuhalten, da eine CVR-Kamera keine Richtung hat. Bei zwei Kamerapositionen gibt es immer Richtungen, die diese Regel verletzen (Abbildung 4-1). Eine zweite Möglichkeit für die Aufnahme von Bildern, die sich ausreichend unterscheiden, ist ein genügend großer Abstand der beiden Kamerapositionen (Beller und Wulff, 2012).

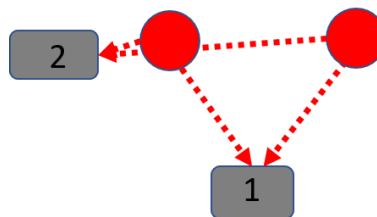


Abbildung 4-1: Die beiden Kamerapositionen (rote Kreise) befolgen die 30°-Regel für Objekt 1, jedoch nicht für Objekt 2.

Die *180°-Regel*, auch *Achsensprung-Regel* genannt, (Bordwell und Thompson, 2013; Kachkovski u. a., 2019) hängt mit Problemen zweiter Art zusammen. Befindet sich eine Aktionsachse in der Szenerie, z. B. vorgegeben durch zwei sprechende Personen, kann es den Betrachter desorientieren, wenn die Kamera diese Linie überspringt. Dieses Thema wird für CVR in **Studie 7.3.6** untersucht und in Abschnitt 4.2.3.4 diskutiert.

Die Reaktion der Zuschauer auf Fehler dritter Art wird in traditionellen Filmen und CVR sehr ähnlich sein: Wenn ein Objekt unmotiviert fehlt, welches in der vorhergehenden Einstellung vorhanden war, ist der Betrachter verwirrt. Es sind allerdings Untersuchungen notwendig, ob es äquivalente Prinzipien für die 30°- und 180°-Regel in CVR gibt.

Bei der Untersuchung verschiedener Schnitttypen mittels Eyetrackers, fanden d'Ydewalle et al. (d'Ydewalle u. a., 1998) heraus, dass die Augenbewegungen bei Schnitten mit Fehlern zweiter und dritter Art nach dem Schnitt zunehmen (im Intervall von 200 – 400 ms). Zu ähnlichen Resultaten kamen Swenberg und Eriksson (Swenberg und Eriksson, 2018). Sie änderten Schnitte so, dass Diskontinuitäten auftraten und stellten fest, dass sich die Sakkaden-Frequenz und die relative Pupillengröße nach solchen Schnitten verstärken. Eine Ursache für die Zunahme der Sakkadenhäufigkeit ist eine visuelle Suche nach dem Eintreten eines neuen visuellen Ereignisses (Smith, 2012).

Montage versus Mis-en-scène

Buccalo (Buccalo, 1977) verglich Mis-en-scène (3.1.1.1) und Montage miteinander, untersuchte deren Wirkung auf die Zuschauer und stellte beide Stile als Endpunkte eines Kontinuums dar (Abbildung 4-2).

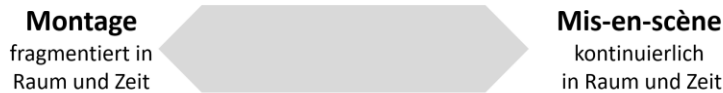


Abbildung 4-2: Kontinuum zwischen Montage und Mis-en-Scène (Buccalo, 1977).

Seine Studien zeigten, dass die Zuschauer die beiden Stile unterschiedlich wahrnehmen und unterschiedliche Präferenzen haben. Trotzdem fühlten sich die Zuschauer bei der Montage näher am Geschehen, empfanden die Handlung realistischer und waren mehr involviert. Viele der heutigen framed Filme befinden sich auf der linken Seite des Kontinuums. Die Entwicklung der Kameratechnik hin zu leichteren Geräten mit hoher Speicherkapazität ermöglichen aber auch Filme aus einer einzigen Einstellung wie „Victoria“ (Schipper 2015), die am rechten Ende des Kontinuums verankert sind.

Für das Verhältnis zwischen Montage und Mis-en-scène in CVR sind neue Analysen notwendig. In unserer **Studie 7.2.5** bevorzugten die Teilnehmer die Variante ohne Schnitte, was zum einen an den gewählten Techniken und Parametern wie auch an den bisherigen Sehgewohnheiten liegen kann. Trotzdem würden die Zuschauer längere Filme mit Schnitte bevorzugen. Noch ist CVR für die meisten Zuschauer eine neue Erfahrung, die im Gegensatz zu anderen Medien der Realität sehr nahekommt, in der es auch keine von außen initiierten Schnitte gibt. Ereignisse, die einem Schnitt gleichkommen, wie das Zuschlagen eines Buches oder eine schnelle weite Augenbewegung sind selbst initiiert. Möglicherweise braucht diese Erkenntnis eine äquivalente Umsetzung in CVR.

Schnitt in CVR

Traditionell wird ein Film auf einer Timeline arrangiert. Beginn und Ende einer Aufnahme werden durch *in-* und *out-Points* bestimmt. Diese Begriffe wurden von Brillhart an CVR angepasst: „the in-point is where a visitor’s experience is most likely to begin and the out-point is where it’s most likely to end“ (Brillhart, 2016). Die *out-Points* können durch das Wissen bestimmt werden, welche Details die Aufmerksamkeit des Betrachters auf sich ziehen, z. B. Bewegungen oder Menschen. Sie können in diesen Regionen platziert und mit der nächsten Aufnahme so verknüpft werden, dass nach dem Schnitt ein für die Story relevantes Detail sichtbar ist. Diese Technik wird *match-in-attention* (Brillhart, 2016) oder *fixedOrientation*-Schnitt (Pavel u. a., 2017) genannt. Es kann jedoch nicht davon ausgegangen werden, dass der Betrachter zum Zeitpunkt des Schnittes wirklich auf den *out-Point* blickt. Daher wird der *in-Point* möglicherweise nicht gesehen.

Werden stattdessen *viewpoint-orientierte* Schnitte (Pavel u. a., 2017) verwendet, kann der Betrachter zuverlässig auf eine Region pro Einstellung aufmerksam gemacht werden: Bei einem *viewpoint-orientierten* Schnitt, wird die nächste Einstellung so ausgerichtet, dass der Betrachter die vordefinierte Rolle sieht, unabhängig von der vorherigen Blickrichtung. Solche Schnitte sind nicht mehr mit herkömmlichen Filmschnittprogrammen und Playern umsetzbar, sondern benötigen eine Implementierung (z. B. mit Unity) und das Anschauen über eine CVR-Applikation.

Fearghail et al. (Fearghail u. a., 2018) analysierten Schnitte und Übergänge in mehreren professionellen CVR-Filmen. In den dabei verwendeten Filmen des Datensatzes von Knorr et al. (Knorr u. a., 2018) sind alle Schnitte *dis-located* und den Zuschauern wurde nach dem Schnitt eine neue Szene präsentiert. Auf diese Weise ändert sich der Hintergrund vollständig. In unseren Untersuchungen konzentrieren wir uns auf Links zwischen *co-located* Kamerapositionen.

4.1.3 Interaktive Narration

Bereits der russische Filmregisseur Sergei Eisenstein hatte den Wunsch nach nichtlinearen Filmen und Büchern, in denen die Geschichte in alle Richtungen weitergehen kann (Tikka, 2004). Mit dem Präsentieren von sphärischen Filmen über HMDs sind wir dem Traum von sphärischen Dramaturgien einen Schritt nähergekommen. Die zusätzliche Raumkomponente bietet Möglichkeiten für natürliche Interaktivität und nichtlineares Storytelling.

Interaktive Filme und VR-Anwendungen

Seit den 1960er-Jahren experimentieren Filmemacher und Forscher mit den Interaktionsmöglichkeiten von Filmen auf Kinoleinwänden und verschiedenen Arten von Displays (Roth, 2015). Interaktive filmische Erzählungen im Web verwenden Spielelemente wie das Lösen von Aufgaben, das Sammeln von Punkten und das Geben von Feedback (Argyriou u. a., 2016). Die Interaktion ist eng mit Engagement verbunden: Die Zuschauer ändern ihre *lean-back* Haltung in eine *lean-forward* Haltung. Vosmeer und Schouten führten für CVR den Begriff *lean-in* Medium ein (Vosmeer und Schouten, 2014): In *lean-in* Medien wird die Blickrichtung und damit auch der Anteil der Information frei gewählt. Der Benutzer ist dadurch aktiver als in klassischen *lean-back* Medien, bleibt aber Zuschauer.

Mehrere VR-Videoplattformen und -Tools gingen in den letzten Jahren online, z. B. WondaVR (Wonda VR, 2019), Panotours (Panotour, 2018), Headjack (Headjack, 2018), Viond (Viond, 2018), InstaVR (InstaVR, 2018). Der Schwerpunkt liegt bei diesen Anbietern auf der Bereitstellung leicht bedienbarer Tools zum Erstellen von Apps für verschiedene VR-Geräte. Zusätzliche Funktionen sind interaktive Elemente für Texte, Bilder, Audios und Szenenlinks. In keinem der Tools existieren jedoch unterschiedliche Designs für Interaktionsregionen oder eine Möglichkeit zum Hinweisen auf Elemente, die außerhalb des Displays liegen. In unserer Arbeit diskutieren wir, welche Arten von Indikatoren geeignet sind, um auf interaktive Elemente aufmerksam zu machen und beschreiben ein Konzept für interaktive CVR-Anwendungen.

Narrative Modelle

Sylvester (Sylvester, 2013) analysiert Game Narratives und unterscheidet zwischen drei Hauptklassen von Stories: *scripted Stories*, *world Narratives* und *emergent Stories*.

Eine *scripted Story* wird von einem Autor festgelegt, ist linear und für jeden Nutzer gleich. Der Autor hat dabei die volle Kontrolle über die Handlung. Reine *scripted Stories* sind in CVR schwer zu realisieren, da die Wahl der Blickrichtung und damit des Sichtfelds eine für CVR wesentliche Form der Interaktion darstellt und der Betrachter die Bildinhalte automatisch beeinflusst. Eine weniger kontrollierende Abwandlung von *scripted Story* ist die *soft-scripted Story*, bei der vorautorisierte Ereignisse abgespielt werden, ohne die Benutzeroberfläche des Nutzers (in CVR das HMD) zu deaktivieren (Sylvester, 2013). Viele aktuelle CVR-Filme fallen in diese Kategorie.

World Narratives erzählen die Geschichte eines Ortes und seiner Menschen. Sie werden über die Konstruktion des Platzes und der darin enthaltenen Objekte/Subjekte erzählt (Sylvester, 2013). *World Narratives* nutzen das Vorhandensein oder Nichtvorhandensein von Merkmalen in der Story World (Sylvester, 2013). Zum Beispiel können kulturelle Symbole, Dokumente oder Töne verwendet werden, um Details zu erzählen. Beispiele für solche Art von CVR-Geschichten sind die Filme „Nomads“ von Felix & Paul Studios (Felix & Paul Studios, 2016).

Emergent Stories sind in vielerlei Hinsicht das Gegenteil von *scripted Stories* (Chauvin u. a., 2014): Sie werden vom Spieler/Benutzer über Interaktionen generiert, z. B. in interaktiven Multiplayer-Games. Der Autor hat wenig Einfluss auf den Verlauf der Geschichte. Er bestimmt das Design von Objekten und Regeln, der Content (Inhalt der Geschichte) wird vom Benutzer erzeugt.

Story-Strukturen

Für die Realisierung interaktiver Geschichten können unterschiedliche Story-Strukturen verwendet werden (Tanak, 2015; Verdugo u. a., 2011). Die einfachste von ihnen, die lineare Struktur, kann auch ohne Interaktion auf einer Zeitachse implementiert werden und entspricht dem traditionellen Film (Abbildung

4-3). Aber auch eine lineare Struktur kann interaktiv implementiert werden. Abbildung 4-4 zeigt ein von uns umgesetztes Beispiel: einen Film aus der Perspektive eines Rollstuhlfahrers. An bestimmten Positionen (z. B. vor einer Treppe) wird der Film erst dann fortgesetzt, wenn der Betrachter das rollstuhlgerechte Hilfsmittel entdeckt hat. In jeder Einstellung ist genau eine Region definiert, die dieses Hilfsmittel enthält (Rampe, Türöffner, Aufzug usw.). Im Gegensatz zu einem herkömmlichen Film wird das Tempo des Films aber durch die Interaktion des Zuschauers bestimmt.



Abbildung 4-3: Lineare Story-Struktur (Sylvester, 2013).

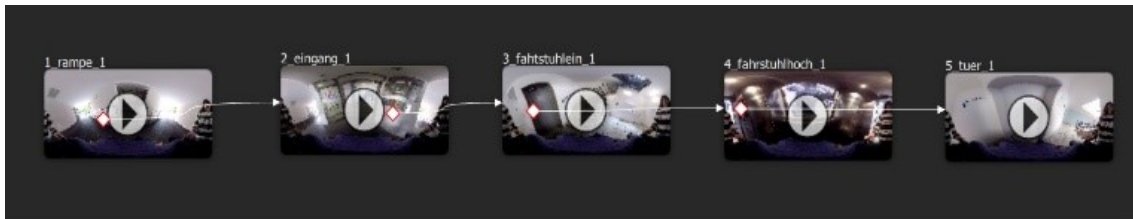


Abbildung 4-4: Beispiel einer interaktiven CVR-Anwendung mit linearer Story-Struktur. Erst wenn der Betrachter einen bestimmten Bereich im sphärischen Filmsegment anschaut, wird der Film fortgesetzt (Darstellung eines eigenen Projektes).

Durch die Erweiterung der Interaktionsmöglichkeiten um weitere Regionen werden nichtlineare Storylines möglich. Das Beispiel in Abbildung 4-5 zeigt ein keltisches Museumsdorf mit mehreren Hütten. In der Eingangsszene (in der Mitte des Bildes) gibt es mehrere Hütten. Für jede der Hütten ist eine Region definiert, die mit der nächsten Szene (innerhalb der Hütte) verbunden ist. Schaut der Zuschauer auf solch eine Region, wird der Film mit einer Szene in dieser Hütte fortgesetzt. Die folgende Szene zeigt einen Druiden, der etwas über das Leben der Kelten erzählt – je nach Hütte zu verschiedenen Themen. In den Hüttensequenzen ist jeweils eine Region an den Türen definiert, durch die der Betrachter in die Dorfszene zurückkehrt. Für die Erzählung ist das Wahrnehmen der Hütten und des Druiden von Bedeutung. Zusätzlich kann sich der Betrachter in den Hütten umsehen und andere Dinge entdecken und Zusatzinformationen erhalten.

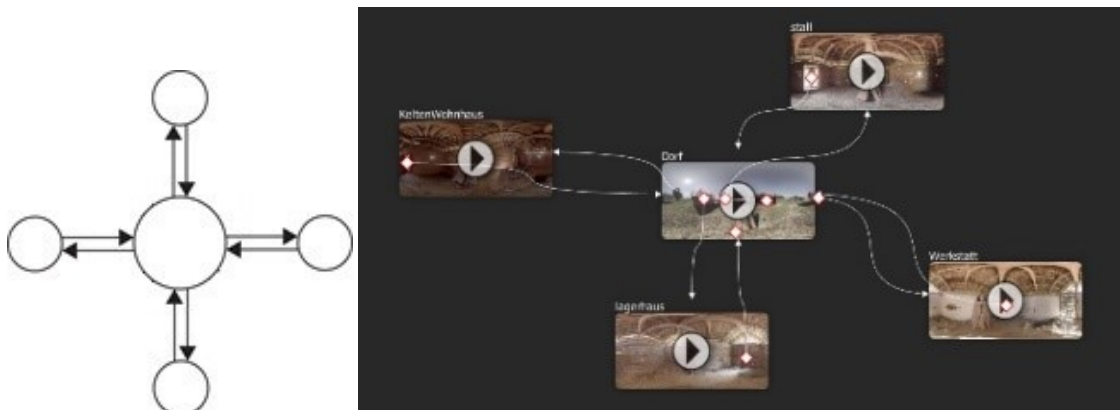


Abbildung 4-5: Links: Beispiel einer nichtlinearen Story-Struktur Rechts: Umsetzung dieser Struktur (Darstellung eines eigenen Projektes).

4.1.4 Interaktion

4.1.4.1 Begriffe

Der Begriff *Interaktion* wird in verschiedenen Gebieten mit unterschiedlichen Bedeutungen verwendet. Interaktionen beschreiben Vorgänge, in denen zwei oder mehr Personen oder Dinge miteinander kommunizieren oder aufeinander reagieren (Cambridge Dictionary, 2019). In diesem Kapitel werden Interaktionen zwischen Menschen und Computern betrachtet, die über ein HMD oder andere technische Geräte realisiert werden. Die Auswahl des Bildausschnitts durch Umherschauen ist eine sehr natürliche Art der Interaktion: Die Kopfbewegungen des Nutzers werden durch die Sensoren des HMDs analysiert und der entsprechende Bildausschnitt wird wiedergeben, so wie es der Nutzer von der realen Welt gewohnt ist.

Interaktionstechniken werden in *Navigation* (Änderung des Ortes), *Selektion* (Auswahl von Objekten), *Manipulation* (Änderung von Objekten) eingeteilt (Bowman u. a., 2001). In CVR kann der Benutzer weder gehen noch Objekte bewegen oder modulieren. Der Inhalt von CVR wird im Gegensatz zu anderen VR Gebieten, bei denen der Inhalt in Echtzeit generiert wird, von Kameras im Voraus erstellt. Die Hauptinteraktionen in CVR sind das Umherschauen und die Selektion von Bereichen.

Für einige CVR-Experiences ist das Umherschauen ausreichend und es sind keine zusätzlichen Interaktionstechniken erforderlich. Die Verwendung von Filmen ist jedoch breit gefächert: Unterhaltung, Dokumentation, Nachrichten, Kunst, Bildung, Marketing (Kandorfer, 2003) ebenso wie Anleitungen und Tutorials. Derzeit sind die meisten CVR-Filme linear, auch wenn die zusätzliche Raumkomponente für nichtlineare Story-Strukturen prädestiniert ist. Interaktive Szenenwechsel in Abhängigkeit von der Blickrichtung können eine natürliche Interaktion sein, ohne die Freiheiten des Betrachters einzuschränken. Dafür bedarf es Selektionstechniken, mit denen der Zuschauer bewusst oder unbewusst Regionen auswählen kann. In dieser Arbeit werden solche zu aktivierende Zielregionen als *Target* bezeichnet.

4.1.4.2 Selektionstechniken in VR

Zur Selektion von Objekten in VR-Umgebungen wurden in den letzten Jahren verschiedene Untersuchungen durchgeführt, von denen einige in diesem Abschnitt näher beschrieben werden. Die für CVR relevanten Schlussfolgerungen werden in Abschnitt 4.2.2.1 erläutert und in einer Tabelle dargestellt.

Für Selektionsprozesse sind unter anderem Kopf-, Augenbewegungen, Gesten und sogar Sensordaten nutzbar. Die meisten dieser Methoden wurden für VR-Umgebungen mit 6 DoFs entwickelt und erfordern eine genaue Prüfung, ob sie auch für CVR geeignet sind (3 DoFs). Einige der Techniken konzentrieren sich auf die Ausführung von Aufgaben oder die Genauigkeit beim Anvisieren eines Zieles. Solche Kriterien sind für CVR oft weniger wichtig, da im Vordergrund eine angenehme Filmerfahrung steht.

Nukarinen et al. (Nukarinen u. a., 2018) verglichen Raycasting (Verwendung eines vom Controller ausgehenden Strahles als Pointer) mit zwei blickbasierten Selektionstechniken. In ihren Experimenten folgte der Strahl der Raycasting Methode der Controller-Richtung und durch Drücken der Controller-Taste wurde das Ereignis ausgelöst. Die beiden blickbasierten Methoden unterschieden sich in der Auslöse-Technik, wobei eine Methode durch die Controller-Taste, die andere durch die Dwell-Zeit (Verweildauer) ausgelöst wurde. Der Strahl erzielte in der Studie bessere Ergebnisse als die Blickmethoden und der Blick bessere als die Dwell-Zeit.

In den Experimenten von Hansen et al. (Hansen u. a., 2018) wurden verschiedene Selektionstechniken bezüglich der benötigten Zeit miteinander verglichen. Die Auswahl mit der Maus erfolgte schneller als mit Blick oder Kopf. Die Dwell-Zeit-Methode war schneller als das Benutzen einer Taste, und die subjektiv empfundene körperliche Belastung fiel für die Selektion mittels Blickes geringer aus als mittels Kopfbewegung. Kallioniemi et al. (Kallioniemi u. a., 2018) untersuchten unterschiedliche Selektions-Arten für Hotspots: Dwell-Zeit und sofortiges Einblenden. Das sofortige Einblenden der nächsten Szene wurde automatisch gestartet, wenn der Kopf in Richtung eines Hotspots zeigte. In ihren Experimenten schnitt

das Einblenden mit großen Symbolen am besten ab. Qian und Teather (Qian und Teather, 2017) verglichen blick- und kopfbasierte Techniken. Die kopfbasierte Selektion war schneller, genauer und bot ein besseres Benutzererlebnis. Auch Choe et al. (Choe u. a., 2018) nutzen in ihren Experimenten die Kopfrichtung zum Adressieren des Targets. Sie verglichen Dwell- und Tastenauslösung für unterschiedliche Dwell-Zeiten und unterschiedliche Targets.

Ein neuartiges Interaktionskonzept, FaceTouch, wurde von Gugenheimer et al. (Gugenheimer u. a., 2016a) entwickelt. Es verwendet die Vorderseite des HMDs als Touchpad, um Bereiche auszuwählen und zu aktivieren. Dies hat den Vorteil, dass auch Elemente aktiviert werden können, die nicht durch die Kopfrichtung ausgewählt sind. Eine weitere neuartige Technik ist Gaze+Pinch (Pfeuffer u. a., 2017). Beide Methoden liefern Eingaben ohne zusätzliche Controller und bieten ein neues Gefühl der Interaktion.

Ein- und Ausgabegeräte sind wichtige Komponenten von Interaktionstechniken. Die Art der erforderlichen Interaktion beeinflusst die Wahl der Methode und des Gerätes. Eingabegeräte können in Geräte mit kontinuierlicher Eingabe (z. B. Tracking) und Geräte mit diskreter Eingabe (z. B. Drücken einer Taste) eingeteilt werden (Bowman u. a., 2001). Kontinuierliche Eingabegeräte sind für Modulationen oder Pointerbewegungen geeignet. Diskrete Eingabegeräte signalisieren ein Ereignis, um beispielsweise eine Auswahl zu bestätigen. Viele Geräte kombinieren beides. Beispielsweise ist eine Mausbewegung kontinuierlich und das Drücken der Maustaste diskret. In verschiedenen VR-Anwendungen ist es wünschenswert, keine weiteren Eingabegeräte neben dem HMD zu verwenden. Mehrere neue Techniken adressieren das Problem handfreier Interaktionsmethoden. Dabei wurden verschiedene Gesten untersucht (Argelaguet und Andujar, 2013; Choumane u. a., 2010; Yu u. a., 2017) oder Pusten als nonverbale Eingabemethode genutzt (Zielasko u. a., 2015).

Es gibt verschiedene Möglichkeiten, Augenbewegungen in der Mensch-Computer-Kommunikation zu nutzen. Jacob et al. (Jacob, 1991, 1990) untersuchten, blickbasierte Interaktionstechniken, welche das Risiko in sich tragen, Ereignisse versehentlich auszulösen, z. B. wenn ein Betrachter einen Bereich längere Zeit inspiziert oder versehentlich in einer bestimmten Weise blinzelt. Dieser Effekt – eine unerwünschte Aktivierung von Befehlen – wird von Jacob als *Midas Touch* beschrieben (Jacob, 1991, 1990). In der griechischen Mythologie wünschte sich König Midas vom Gott Dionysos, dass alles, was er berührt, zu Gold werden solle. Er konnte dadurch neue Schätze schaffen, allerdings nicht mehr essen und trinken.

Mehrere Studien befassen sich mit der Vermeidung solcher versehentlich ausgelösten Aktionen (Mohan u. a., 2018; Velichkovsky u. a., 1997). Mardanbegi et al. (Mardanbegi u. a., 2012) führten blickbasierte Kopfgesten ein und kombinierten Kopf und Blickbewegungen so, dass der Blick des Benutzers den Fokus auf das Interaktionsobjekt behält. In ihren Experimenten verwendeten sie einen am Kopf befestigten Eyetracker. Der Betrachter fixiert das Objekt und bewegt den Kopf auf und ab. Diese Methode wurde für flache Displays und Fahrzeugschnittstellen evaluiert (Mardanbegi und Hansen, 2012). Die Ergebnisse zeigen die Anwendbarkeit dieser Methode für die Interaktion in diesen Bereichen. Schnelle Kopfbewegungen gehen mit dem vestibulookulären Reflex (VOR) einher (Fetter, 2007), bei dem die Augen mit der gleichen Geschwindigkeit in die entgegengesetzte Richtung bewegt werden, damit ein Objekt noch fixiert werden kann. Dieser Reflex unterstützt die Stabilisierung des Bildes auf der Netzhaut und wird für blickbasierte Kopfgesten ausgenutzt. Špakov und Majaranta (Špakov und Majaranta, 2012) fünf verschiedene blickbasierte Kopfgesten: Nicken, Drehen nach links / rechts und Neigen nach links / rechts. Die Teilnehmer empfanden das Nicken als geeignetste Technik für Selektionsaufgaben.

4.2 Spaceline-Konzept

Löst man sich von den Werkzeugen der traditionellen Filmproduktion und nutzt die Möglichkeiten, die VR bietet, eröffnen sich Optionen, um neue Elemente für das Storytelling einzusetzen. Da in CVR neben der Zeit der Raum als weitere Komponente nutzbar ist, ist es überlegenswert, Schnitte nicht nur von der verstrichenen Zeit abhängig zu machen, sondern auch von der Blickrichtung des Zuschauers. Der Filmemacher erhält damit eine weitere Möglichkeit, Bildwechsel herbeizuführen. In Anlehnung an den Begriff Timeline, führen wir dazu den Begriff *Spaceline* ein. Für ein *Spaceline-Konstrukt* (Abschnitt 4.2.1)

werden verschiedene Regionen festgelegt: zum einen Regionen, die einen Szenenwechsel bewirken und zum anderen Regionen, die der Zuschauer nach dem Szenenwechsel als Erstes sieht. Dazu braucht es neben diesen Regionen Selektionsmethoden, die den Szenenwechsel auslösen (Abschnitt 4.2.2) und Techniken für den Übergang zwischen den Szenen (Abschnitt 4.2.3).

4.2.1 Terminologie

4.2.1.1 Einstellung, Szene, Transition und Space

Zwei Grundbegriffe der traditionellen Filmmontage sind *Einstellung* und *Szene*. Während eine *Einstellung* ein Filmsegment zwischen zwei Schnitten beschreibt (Bender u. a., 2012), repräsentiert eine *Szene* eine Einheit eines Films mit Kontinuität in Zeit, Ort und Charakteren (Wulff, 2012e). Sie besteht meist aus mehreren Einstellungen. In CVR ist die Anzahl der Schnitte geringer, da der Betrachter selbst verschiedene Bildausschnitte in der Szenerie auswählt. Oft hat eine Szene keine weiteren Schnitte. Der Übergang zwischen den Szenen wird *Transition* genannt (Wulff, 2012f).

In framed Filmen stimmen das Bild der Kamera und das des Zuschauers überein. In CVR gibt es zwei Perspektiven: die Rundumsicht der Kamera und das kleinere, selbst gewählte Sichtfeld (FoV) des Betrachters. Der Begriff *Einstellung* ist daher nicht direkt übertragbar, es werden zwei Begriffe für das Filmsegment zwischen zwei Schnitten benötigt. Wir definieren daher einen *Space*: Ein **Space** ist ein sphärisches Filmsegment, das ohne Unterbrechung aufgezeichnet wurde. Die **Einstellung** ist die Bildsequenz, die der Betrachter im FoV in diesem Space ausgewählt hat. Diese ist nicht sphärisch, sondern entspricht dem Blickfeld des Betrachters im HMD. Als **Spaceline** bezeichnen wir einen Pfad durch eine Struktur von Spaces. Diese Struktur wird vom Filmemacher entworfen und heißt **Spaceline-Konstrukt**. Darin bestimmt jeder Betrachter seine Spaceline – einen Pfad durch dieses aus mehreren Spaces bestehenden Konstruktes. Im Gegensatz zu einem timeline-basierten Film, der allein vom Filmemacher bestimmt wird, wird die Spaceline vom Filmemacher und vom Betrachter bestimmt.

Beim traditionellen Filmemachen wie auch in vielen CVR-Applikationen wird ein Schnitt auf einer Timeline definiert. Die nächste Einstellung startet nach einem festgelegten Zeitintervall. Wir nennen solche Schnitte **timeline-basiert**. Die zusätzliche Raumkomponente in CVR bietet die Möglichkeit, den Raum als Referenz zu verwenden: Die nächste Einstellung beginnt, wenn der Betrachter in einen bestimmten Bereich schaut. Solche Übergänge nennen wir **spaceline-basierte Transitions**. Für beide Arten von Übergängen (time, space) kann in CVR die Blickrichtung für die nächste Einstellung ausgerichtet werden. In den **Studien 7.3.1 und 7.3.6** wurden verschiedene Ausrichtungsmethoden untersucht, um herauszufinden, in welcher Weise sie eingesetzt werden können.

4.2.1.2 Regionen

Das Spaceline-Konzept beinhaltet verschiedene Arten von Regionen: Eine **out-Region** ist die Region, deren Aktivierung eine Einstellung beendet. Von dort erfolgt der Wechsel über eine *Transition* zum nächsten Space, in welchem der Betrachter zuerst die *in-Region* sieht, und von der aus die Szenerie dann erkundet werden kann. Eine **in-Region** ist dabei der Bereich, den der Zuschauer zu Beginn einer Einstellung sieht.

Darüber hinaus führen wir **add-Regionen** ein, die dem Betrachter zusätzliche Interaktionsoptionen bieten, z. B. das Vergrößern von Details oder das Abrufen zusätzlicher Textinformationen.

Das vom Filmemacher festgelegte Spaceline-Konstrukt verbindet *out-Regionen* mit *in-Regionen* durch **Links**. Auf diese Weise werden vom Betrachter ausgelöste Space-Wechsel interaktiv. Für *nichtlineare* Storys kann mehr als eine *out-Region* definiert werden. Der Wechsel des Spaces und damit auch der Einstellungen erfolgt durch eine **Transition**, in dem Moment in dem der Zuschauer die *out-Region* bewusst oder auch unbewusst aktiviert.

Abbildung 4-6 illustriert ein Spaceline-Konstrukt an einem Beispiel:

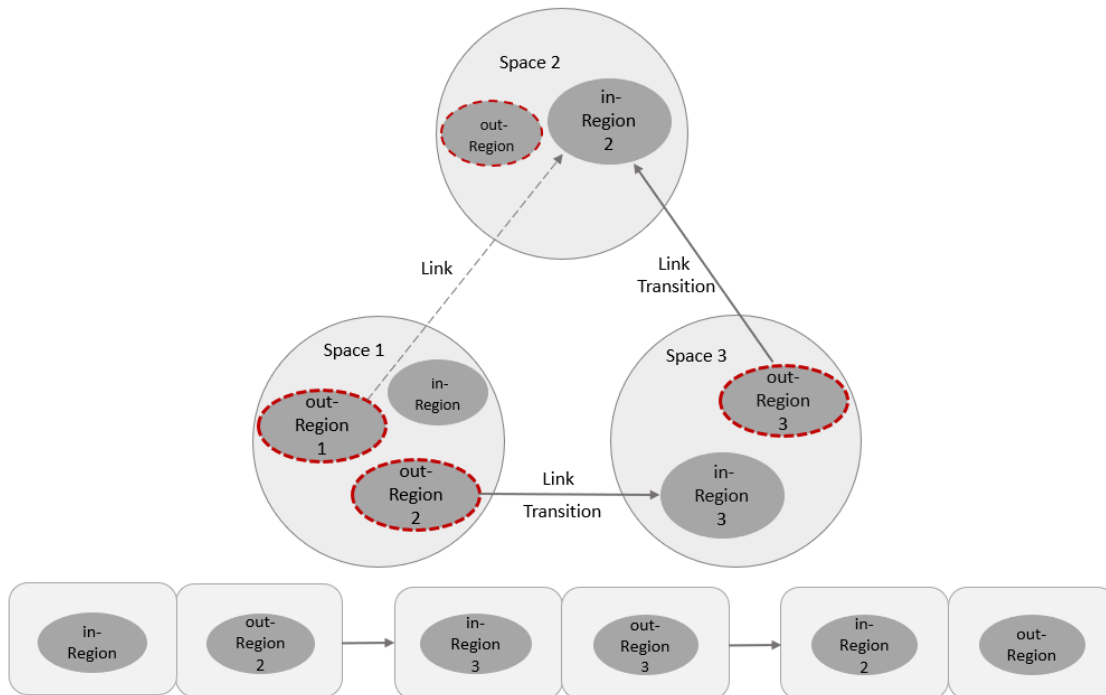


Abbildung 4-6: Grafische Darstellung eines Beispiels für ein Spaceline-Konstrukt aus drei Spaces. Jeder Space hat eine *in-Region*, in die der Zuschauer als Erstes sieht. Space 1 hat zwei *out-Region*, die mit unterschiedlichen Spaces verlinkt sind. Wird auf *out-Region 2* geschaut, befindet sich der Zuschauer nach einer Transition in Space 3, die über die *out-Region 3* verlassen werden kann, um in Space 2 zu gelangen. Im unteren Teil der Darstellung sind die dazugehörigen Einstellungen im FoV des Zuschauers dargestellt.

4.2.1.3 Indikatoren

Um dem Betrachter das Erkennen von Regionen und die damit verbundenen Interaktionsmöglichkeiten zu erleichtern, führen wir verschiedene Visualisierungsmethoden ein. Diese sind spezielle Methoden zur Aufmerksamkeitslenkung und wurden in Kapitel 2 ausführlich behandelt. Es ist wichtig, dass diese Visualisierungen das Seherlebnis möglichst wenig beeinträchtigen. Dabei unterscheiden wir zwei Fälle: Zum einen können Bereiche im aktuellen Sichtfeld des Betrachters liegen und durch *on-screen Indikatoren* kenntlich gemacht werden (Abbildung 4-7, links). Andererseits weisen *off-screen Indikatoren* auf Bereiche außerhalb des Displays hin, um dem Betrachter das Auffinden zu erleichtern (Abbildung 4-7, rechts).



Abbildung 4-7: Verschiedene Beispiele von on-screen (links) und off-screen (rechts) Indikatoren. Von links nach rechts: (1) Die Region wird mit einer transparenten farbigen Fläche überlagert. (2) Ein Pfeil zeigt auf die Region. (3) Am linken Displayrand weist ein Indikator auf eine Region links vom FoV hin. (4) Ein Strahl verbindet die Mitte des FoVs mit einer Region.

Für die verschiedenen Medien gibt es unterschiedliche Techniken, auf Regionen außerhalb des Bildschirms hinzuweisen: auf flachen Displays (Gustafson u. a., 2008; Gustafson und Irani, 2007; Hossain u. a., 2012; Zellweger u. a., 2003), in 3D Umgebungen (Jo u. a., 2011) und in AR (Gruenefeld u. a., 2017b, 2017a). Indikatoren sind spezielle Methoden der Aufmerksamkeitslenkung und wurden in Kapitel 2 ausführlich beschrieben.

4.2.1.4 Selektion, Pointer, Trigger

Das Spaceline-Konzept benötigt zusätzlich Techniken, um die Auswahl und Aktivierung von Rols zu ermöglichen. Selektionstechniken umfassen zwei Phasen: Als Erstes wird die Richtung ausgewählt, in der sich das Target befindet. Wir nennen diesen Prozess **Pointing**. Das Ereignis, welches die Auswahl auslöst, z. B. der Ablauf der Dwell-Zeit oder ein Nicken, nennen wir **Trigger**, den dazugehörigen Vorgang **Triggern**. Oft sind Pointing und Triggern mit dem gleichen Gerät möglich, zum Beispiel ein Bewegen der Maus und der zugehörige Mausklick.

Für das bewusste oder auch unbewusste Auswählen einer Region durch **Pointing** wird ein **Pointer** benötigt, der die aktuell gültige Position anzeigt. In CVR ist dies in der Regel die Kopf- oder Blickrichtung.

4.2.2 Selektionstechniken in CVR

Für die Realisierung des Spaceline-Konzeptes braucht es Selektionstechniken, um *out*- und *add*-Region auszuwählen. Augen- und Kopfbewegungen sind natürliche Techniken zur Auswahl eines Sichtfeldes und kontinuierliche Eingabemethoden, die zum Adressieren eines Targets geeignet sind. Als Trigger wird ein diskretes Signal benötigt, welches der Kopf durch Nicken initiieren kann, die Augen beispielsweise durch Blinzeln. Auch wenn es sich beim Nicken und Blinzeln um Prozesse handelt, kann ein diskreter Moment definiert werden, der für das Senden des Signals verantwortlich ist. Gleiches gilt für die Dwell-Zeit: nachdem ein Objekt eine definierte Zeit lang betrachtet wurde, wird das Ereignis ausgelöst.

4.2.2.1 Erkenntnisse aus Arbeiten anderer Autoren

Für einen besseren Überblick über die in anderen Arbeiten verwendeten Methoden, werden diese in Tabelle 4-1 zusammengefasst. Dabei ist jede Methode in ihre Pointing- und Trigger-Komponente aufgeteilt. Die Inhalte der Veröffentlichungen wurden in 4.1.4.2 detailliert beschrieben.

Literatur	Pointing	Trigger	Studienergebnis
(Nukarinen u. a., 2018)	Controller Auge	klick dwell	Controller am besten, klick besser als dwell
(Hansen u. a., 2018)	Kopf Auge Maus	dwell (300 ms) klick	dwell schneller als klick, Targetgröße sollte für die blickbasierte Methode mindestens 3° sein, Performance besser für kopfbasierte Methode, blickbasierte Methode weniger anstrengend
(Kallioniemi u. a., 2018)	Kopf	dwell (2 s) prompt (fade-in)	am besten: große Icons, zentrierte Ausrichtung
(Qian und Teather, 2017)	Auge Kopf Auge+Kopf	klick	Kopf am besten
(Choe u. a., 2018)	Kopf	klick dwell	klick wird für allgemeine Aufgaben empfohlen, Targetgröße mehr als 3° 50' Dwell-Zeit von 1 s wird empfohlen
(Gugenheimer u. a., 2016a)	Kopf/Hand	touch	Neue Methode: FaceTouch
(Pfeuffer u. a., 2017)	Auge	Gesten	Neue Methode: Gaze+Pinch

(Zielasko u. a., 2015)	Controller	pusten	Neue Methode: BlowClick
(Mardanbegi u. a., 2012)	Auge	Kopf: nicken	Neue Methode: blickbasierte Kopfgesten
(Špakov und Majaranta, 2012)	Auge	Kopf: nicken, drehen, neigen	nicken - am besten für Selektion drehen - für Navigation neigen – für Funktionsumschaltungen

Tabelle 4-1: Veröffentlichte Selektionsmethoden für VR (*click* = Betätigen einer Controller-Taste).

Es gibt eine Reihe von Forschungsergebnissen, die für das Spaceline-Konzept relevant sind:

Kopfbasierte Techniken

Kopfbasierte Techniken haben den Vorteil, dass keine zusätzlichen Geräte benötigt werden und alle Instrumente (Tracker) bereits im HMD verfügbar sind. Für Aufgaben, bei denen der Benutzer schnell zwischen weit voneinander entfernten Bereichen wechseln muss, kann die intensive Verwendung kopfbasierter Pointings jedoch anstrengend sein (Hansen u. a., 2018). Kopfbasierte Pointing-Techniken eignen sich gut für CVR, da hier die Auswahl nicht schnell sein muss und nur wenige Bereiche benötigt werden.

Blickbasierte Techniken

Mit blickbasierten Techniken können alle auf dem Display sichtbaren Objekte schnell adressiert und getriggert werden. Die Augen bewegen sich jedoch ungleichmäßiger als der Kopf und das Pointing ist weniger genau (Hansen u. a., 2018). Dies kann teilweise durch den verwendeten Eyetracker oder auch durch eine ungenaue Kalibrierung verursacht sein. Blickbasierte Trigger-Techniken können anstrengend sein, wenn sie häufig und lange angewendet werden.

Dwell-Zeit

Für kopf- und blickbasierte Techniken kann die Dwell-Zeit für die Triggerphase genutzt werden. Kurze Dwell-Zeiten können einen Midas Touch verursachen und die Fehlerrate erhöhen (Jacob, 1990). Zu lange Dwell-Zeiten verlangsamen die Taskperformance und können auf den Benutzer störend wirken.

Es gibt mehrere Studien über die Dauer der Dwell-Zeit. Die Ergebnisse hängen von der untersuchten Aufgabe und dem benutzten Gerät ab. Choe et al. (Choe u. a., 2018) untersuchten in ihren Studien mehrere Dwell-Zeiten (1 s, 1.5 s, 2 s) und empfahlen eine Dwell-Zeit von 1 s für Selektionsaufgaben mittels HMD. Eine Dwell-Zeit von 300 ms wird für einen erfahrenen Benutzer in einer Selektionsaufgabe als machbar angesehen (Hansen u. a., 2018). Kallioniemi et al. (Kallioniemi u. a., 2017) verglichen interaktive VR Filmerlebnisse in einem CAVE-System mit der Verwendung eines HMDs. In ihren Experimenten ermittelten sie eine optimale Dwell-Zeit von 5 s für das CAVE-System und 2 s für die HMD-Anwendung.

Feedback

Majaranta et al. (Majaranta u. a., 2004) untersuchten Feedback und Dwell-Zeit bezüglich der Geschwindigkeit und Genauigkeit der Eingabe. Sie zeigten, dass Feedback das Blickverhalten, die Genauigkeit und die subjektive Erfahrung beeinflusst. Außerdem stellten sie fest, dass kurze Dwell-Zeiten ein kurzes Feedback erfordern.

Target

Wenn ein Targetbereich zu klein ist, ist es schwierig, ihn zu adressieren. Ist er zu groß, kann er einen Midas Touch verursachen. Hansen et al. (Hansen u. a., 2018) empfehlen, dass der Targetbereich für blickbasierte Interaktionen mindestens 3° betragen sollte. Targetgrößen unter 3° sind bei den meisten Eingabemethoden schwierig auszuwählen (Qian und Teather, 2017). Choe et al. (Choe u. a., 2018) verglichen zwei verschiedene Targetgrößen (1°55' und 3°50') und empfehlen eine Targetgröße von mehr als 3°50'.

4.2.2.2 Nutzerstudie: Blickbasierte Interaktionstechniken

Um mehr über die Wirkung von Selektionstechniken in CVR zu erfahren, haben wir in einem ersten Schritt mehrere blick- und kopfbasierte Techniken miteinander verglichen. Auch wenn Gesten natürliche Interaktionen im realen Leben sind, werden sie beim Schauen von Filmen wenig verwendet. Basierend auf diesen Überlegungen wurden in der **Studie 7.3.5** die folgenden vier Trigger-Techniken implementiert: die blickbasierte Kopfgeste (**nod**), die Dwell-Zeit (**dwell**), das Blinzeln (**blink**) und das Drücken einer Controller-Taste während auf die RoI geblickt wird (**click**) (Abbildung 4-8).

- **Nod:** Der Betrachter nickt mit dem Kopf, während die Augen die RoI fixieren.
- **Dwell:** Der Betrachter fixiert die RoI für ein bestimmtes Zeitintervall.
- **Blink:** Der Betrachter fixiert die RoI und blinzelt zweimal kurz hintereinander.
- **Click:** Der Betrachter schaut auf die RoI und drückt die Taste eines Mini-Bluetooth-Controller.

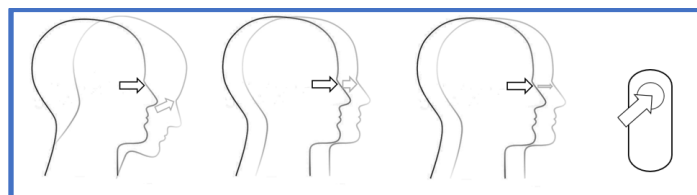


Abbildung 4-8: Die vier Selektionsmethoden, von links nach rechts: *nod*, *dwell*, *blink*, *click*.

Für die meisten Aspekte wurde die *click*-Methode am besten bewertet. Dies kann damit zusammenhängen, dass die Verwendung eines Controllers zur Interaktion generell eine Methode ist, mit der die meisten Menschen aus anderen Szenarien vertraut sind. Die Zuschauer sind an Controller wie Computermouse und Fernbedienungen gewöhnt. Diese Art von Interaktionen ist daher einfach anwendbar. Die Teilnehmer bewerteten die anderen Methoden jedoch als kreativer, erfinderischer und neuartiger.

Mit den kopf- und blickbasierten Methoden waren nicht alle Teilnehmer zufrieden, was mehrere mögliche Ursachen haben kann: Einerseits sind die Schwellwerte der kopf- und blickbasierten Methoden nicht für jede Person gleich und funktionierten deshalb für einige Teilnehmer nicht ausreichend. Auch die Genauigkeit des Eyetrackers kann die Wirksamkeit beeinflussen. Auf der anderen Seite ist davon auszugehen, dass der Betrachter in CVR-Erlebnissen die Augen nicht ständig für eine aktive Interaktion nutzen möchte, da diese bereits mit dem Ansehen des Films beschäftigt sind. Blickbasierte *Pointing* Techniken sind jedoch natürlich, da der Betrachter oft automatisch in die Richtung der RoI schaut.

Blick- und kopfbasierten Methoden hängen von den gewählten Parametern ab: der Winkel für das Nicken, die Dauer für die Dwell-Zeit, die Schwellwerte für das Blinzeln. Unsere Ergebnisse wurden deshalb von den gewählten Werten beeinflusst. Auch wenn wir die Parameter in einem Pilotversuch sorgfältig ausgewählt und die Hinweise der Literatur berücksichtigt haben, funktionierten sie nicht für jeden Studienteilnehmer gleich. Darüber hinaus beeinflusst die Genauigkeit des Eyetrackers die Wirksamkeit der Methoden. Die *click*-Methode, bei deren Aktivierung der Eyetracker nicht benötigt wird, funktionierte effizient. Einige Teilnehmer erwähnten jedoch, dass der Controller die Presence beeinträchtigt. Dies stimmt nicht mit den Ergebnissen des Presence-Fragebogens überein, bei dem der Controller keine signifikanten Unterschiede bezüglich der Presence im Vergleich zu den anderen Methoden aufwies. Vor unserer Studie gingen wir ebenfalls davon aus, dass der Controller für die Presence störend sein könnte, was durch die Ergebnisse des Presence-Fragebogens allerdings nicht bestätigt wurde. Es scheint, als würde ein einfacher Ein-Tasten-Controller als verlängerter Daumen fungieren, und der Betrachter ist sich dessen während der Anwendung nicht wirklich bewusst. Weitere Untersuchungen sind erforderlich, um diese Behauptung zu überprüfen. Bei allen Teilnehmern funktionierte diese Methode korrekt, im Gegensatz zu den blickbasierten Methoden, die von den Parametern und Teilnehmern abhängen. Unsere Studien können nur ein erster Schritt sein, um herauszufinden, welche Methoden und Parameter für CVR-Experiences geeignet sind und wie sie eingesetzt werden können.

4.2.3 Transitions: Montagetechniken in CVR

Um von einem Space in den nächsten zu gelangen, können verschiedene Transition-Techniken genutzt werden. In unserem Spaceline-Konzept beinhalten diese Techniken nicht nur die zeitliche Veränderung der Bildinhalte (z. B. durch Überblendung), sondern auch die Ausrichtung des Blickes des Zuschauers.

4.2.3.1 Erkenntnisse aus Arbeiten anderer Autoren

Um Möglichkeiten für visuelle Übergänge zwischen Szenen in VR zu untersuchen, wurden in den letzten Jahren von mehreren Autoren Transition-Methoden eingeführt und miteinander verglichen (Tabelle 4-2):

Literatur	Umgebung	Methoden
(Men u. a., 2017)	VR	simple cut, super fast, fade, vortex
(Oberdörfer u. a., 2018)	VR	simulated blink, turn around, virtual HMD
(Medeiros u. a., 2016)	VR	teleport, linear motion, animated teleport box
(MacQuarrie und Steed, 2018)	CVR	teleport, lineare Bewegung durch ein 3D-Model der Szene, image-basierte Transition mittels Möbius-Transformation

Tabelle 4-2: Veröffentlichte Transition-Methoden für VR und CVR.

Alle diese Transitions beschreiben imagebasierte Effekte zwischen den Szenen, die oft auch in framed Filmen verwendet werden. In CVR ist für die Bestimmung einer Transition daneben auch die Richtung wichtig, in die der Zuschauer zu Beginn der neuen Szene schaut. Dieser Aspekt wird im nächsten Abschnitt näher untersucht.

4.2.3.2 Nutzerstudie: Analyse von Ausrichtungsmethoden in CVR

Um die erweiterten Möglichkeiten des Spaceline-Konzeptes für co-located Spaces zu nutzen, führen wir neue Transition-Techniken ein, bei denen nicht nur der neue Space spezifiziert wird, sondern auch die initiale Blickrichtung des Betrachters in diesem Space (Abbildung 4-9).

- **toRoI:** Die Blickrichtung zeigt nach der Transition zu einer vorher definierten RoI, unabhängig davon, wohin der Betrachter zuvor geschaut hat. Diese Methode entspricht den viewpoint-orientierten Schnitten von Pavel et al. (Pavel u. a., 2017)
- **keepFocus:** Die Kamera ändert die Position zu einem vom Filmemacher festgelegten Ort, der Fokus des Betrachters bleibt jedoch unberührt. Das Objekt, das sich vor der Transition im Sichtfeld befand, ist auch danach zu sehen
- **keepDir:** Die Blickrichtung wird beibehalten - nur die Position ändert sich.

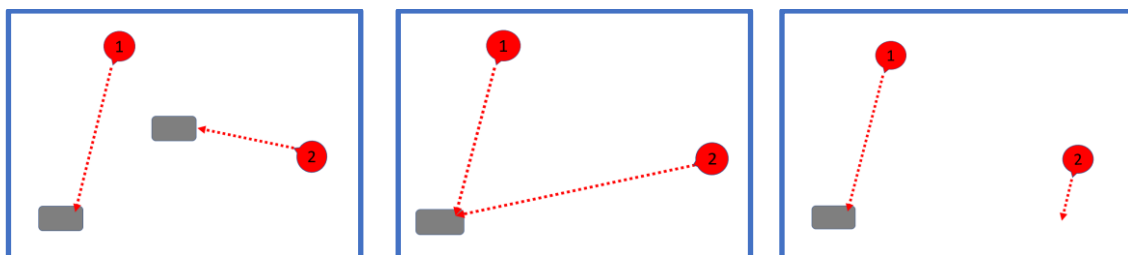


Abbildung 4-9: Illustration der drei Methoden: Die Pfeile zeigen die Blickrichtungen vor (1) und nach (2) der Transition. Links: nach einer *toRoI*-Transition zeigt das FoV auf eine RoI. Mitte: Nach einer *keepFocus*-Transition zeigt die FoV-Richtung auf die vorher inspizierte Region. Rechts: bei einer *keepDir*-Transition ändert sich die Richtung nicht.

In **Studie 7.3.1** untersuchten wir co-located Transitions, mit verschiedenen Ausrichtungen. Die Ergebnisse der Studie haben gezeigt, dass viele Teilnehmer es vorziehen würden, Transitions selbst zu initiieren. Da die Transitions *timeline-basiert* waren, konnte es vorkommen, dass der Betrachter etwas inspizierte und nicht für die Transition bereit war. Außerdem kann eine zeitgesteuerte *keepFocus*-Transition in einem Moment auftreten, in dem der Benutzer nichts fokussiert. In solchen Fällen kann die neue Blickrichtung ein uninteressanter Bereich sein, z. B. eine leere Wand.

Die Methode *toRol*, bei der der Betrachter nach der Transition in Richtung einer definierten Rol schaut, kann sowohl für *timeline-basierte* Transitions als auch für *spaceline-basierte* Transitions, die von der Blickrichtung der Zuschauer abhängen, verwendet werden. Allerdings muss die vertikale Position berücksichtigt werden, um sicherzustellen, dass die Rol im Sichtfeld des Betrachters liegt. Die Methode wirkte gut für auffällige Objekte auf Augenhöhe des Betrachters. Sie kann verwendet werden, um den Benutzer zu wichtigen Rols zu führen. Abhängig von der Geschichte kann dies auch für co-located Aufnahmen funktionieren: Wenn es eine Kontinuität in der Story gibt, bemerken die Zuschauer Diskontinuitäten in Raum und Richtung nicht. Bei Erkundungsszenen mit weniger hervorstechenden Rols oder ohne fortlaufende Story kann die *toRol*-Methode den Betrachter desorientieren, da der Raum inspiziert wird und anders ausgerichtet ist als zuvor. Um dies zu vermeiden, sollte der Bereich, der nach der Transition gesehen wird, von einer Art sein, die die Aufmerksamkeit des Betrachters auf sich zieht.

Die Ergebnisse der **Studie 7.3.1** bekräftigen die Resultate anderer Forschungen (Pavel u. a., 2017), die zeigen, dass *toRol*-Transitions verwendet werden können, um die Aufmerksamkeit in co-located und dis-located Szenarien zu lenken. Co-located Aufnahmen benötigen nach der Transition eine fortlaufende Handlung mit sprechenden Personen oder anderen herausragenden Details in der Rol, die die Aufmerksamkeit des Zuschauers fesseln. Co-located *toRol*-Transitions eignen sich jedoch nicht für Filme mit subtilen Details, die der Betrachter entdecken muss. Außerdem funktioniert die Methode nur, wenn die neue Rol die gleiche Höhe hat wie der Blick des Betrachters. Um den Betrachter auf andere Objekte aufmerksam zu machen, die subtiler, niedriger oder höher sind, haben wir eine weitere neue Methode implementiert und untersucht: die *wait*-Methode (Abschnitt 4.2.3.3).

Die *keepFocus*-Methode, bei der der Betrachter nach einer Transition von einer anderen Position aus in die gleiche Region schaut wie zuvor, sollte nur für *spaceline-basierte* Transitions verwendet werden. *Timeline-basierte* Transitions während einer zufälligen Augenbewegung können den Zuschauer verwirren. Diese Art der Transition sollte nur ausgelöst werden, wenn der Betrachter einen Bereich für eine bestimmte Zeit fixiert oder wenn der Betrachter die Transition auf bewusste Weise auslöst.

Eine Anwendung der *keepFocus*-Technik ist die Verwendung von Einstellungsgrößen (z. B. Nahaufnahme, Halbnah, Totale). Solche Einstellungsgrößen sind wichtige Stilelemente im traditionellen Film, um die Beziehung zum Charakter zu zeigen und Gefühle hervorzurufen. Den Abstand der Kamera zu einem Motiv zu ändern, ist jedoch nur dann sinnvoll, wenn der Betrachter bewusst in diese Richtung blickt. Dadurch kann eine Transition ausgelöst werden und die nächste Kameraposition ist näher am Motiv, die Blickrichtung ist jedoch wieder auf denselben Bereich fokussiert. Es ist zu beachten, dass sich die Bildausschnitte vor und nach der Transition wesentlich voneinander unterscheiden. Bei *spacelinebasierten* Transitions ist die Blickrichtung vor und nach der Transition bekannt, so dass die Position der Kamera so bestimmt werden kann, dass die 30°-Regel eingehalten wird.

4.2.3.3 Nutzerstudie: *wait*-Methode

Mit den obigen Methoden können die Teilnehmer nicht direkt zu Rols geführt werden, die sich nicht auf ihrer aktuellen vertikalen Augenhöhe befinden. Daher sind für solche Bereiche weitere Forschungsarbeiten erforderlich. In **Studie 7.3.2** wurde ein erster Ansatz untersucht. Dafür wurden zwei Transition-Methoden zur Unterstützung der Aufmerksamkeit mit einer Baseline verglichen. Bei der ersten Methode sieht der Betrachter die Rol nach der Transition, unabhängig von der Blickrichtung zuvor, wie bereits in den **Studien 7.3.1 und 7.3.6** (*toRol*). Bei der zweiten Methode wird der Film nicht fortgesetzt, bevor der Betrachter die definierten Rols gesehen hat (*wait*-Technik).

- **toRol:** Die Blickrichtung zeigt nach der Transition zu einer vorher definierten Rol, unabhängig davon, wohin der Betrachter zuvor geschaut hat.
- **wait:** Der Film wird erst dann fortgesetzt, wenn der Betrachter alle Rols einer Szene gesehen hat. Dabei wird eine Rol als „gesehen“ betrachtet, wenn der Betrachter mindestens zwei Sekunden lang in diese Richtung blickte. Nach einem Zeitlimit, wird das Video in jedem Fall fortgesetzt, auch wenn nicht alle Rols gesehen wurden.

Beide Methoden wurden von den Teilnehmern akzeptiert und in unseren Tests der Baseline (ohne Methode) vorgezogen. Für die in der Studie verwendeten Filme wurde die *wait*-Methode bevorzugt. Ein Vorteil dieser Methode ist die geringere Angst, etwas zu verpassen und mehr Zeit für Details zu haben. Für unauffälligere Elemente oder Bereiche, die weiter von der horizontalen Linie entfernt sind, sollte die *wait*-Methode verwendet werden. Diese Methode ist auch nützlich, wenn eine Aufnahme mehr als eine wichtige Rol enthält.

Wir sind uns bewusst, dass unsere Ergebnisse durch die Art des verwendeten Videomaterials beeinflusst sein können. Für Filme mit mehr Action und konzentrierten einzelnen Rols sollte die *toRol*-Methode bevorzugt werden. Es bedarf weiterer Forschungen, um genauer herauszufinden, wie die Filmtypen die Wirkungen der Methoden beeinflussen.

4.2.3.4 Nutzerstudie: Achsensprung-Problematik

Die **Studie 7.3.1** setzt sich mit der 180°-Regel auseinander, die auch als Achsensprungproblem bekannt ist. Dabei wurde der Schwerpunkt auf die Platzierung der Kamera und die Blickrichtung in Bezug auf die 180°-Achse gelegt. Beim traditionellen Filmemachen kann ein Verstoß gegen diese Regel zu einer Desorientierung des Zuschauers führen: Was vorher links war, ist nach der Transition rechts und umgekehrt. Um dies zu vermeiden, schreibt die Regel vor, dass die Kamera nicht auf die andere Seite der Aktionslinie wechseln sollte. Auch wenn diese Regel verletzt werden darf, ist es für jeden Filmemacher wichtig, sie zu verstehen. Die Frage stellt sich, ob es auch in CVR Kamerapositionen gibt, die aufeinanderfolgend den Zuschauer verwirren. Dabei ist zu beachten, dass in CVR auch die Blickrichtung des Betrachters relevant ist. Möglicherweise ist nicht die 180°-Achse das entscheidende Kriterium, um Szenenübergänge so zu gestalten, dass sie beim Betrachter keine Desorientierung hervorrufen. Für die Studie wurden drei Bedingungen miteinander verglichen: zwei verschiedene Kameraausrichtungen (*toRol*, *personView*) und eine Baseline ohne Transitions.

- **toRol:** Die Blickrichtung zeigt nach der Transition zu einer vorher definierten Rol, unabhängig davon, wohin der Betrachter zuvor geschaut hat.
- **personView:** Der Betrachter nimmt die Position einer Person in der virtuellen Welt ein und die Blickrichtung wird an der virtuellen Welt ausgerichtet.

Keiner der getesteten Fälle verstieß im herkömmlichen Sinne gegen die 180°-Regel – die Kamera überquerte nicht die Achse zwischen zwei sprechenden Charakteren. Die *personView*-Methode führte jedoch zu einer Desorientierung des Betrachters. Am Tisch saßen die sprechenden Personen vor der Transition auf der rechten Seite des Betrachters und nach der Transition auf der linken Seite – ähnlich dem 180°-Problem beim traditionellen Filmemachen. Trotzdem scheinen co-located Links möglich zu sein, wenn die Kamera nach der Transition auf eine Rol ausgerichtet wird. In solchen Fällen gibt es eine Kontinuität in der Geschichte und die Diskontinuität im Raum zieht die Aufmerksamkeit des Betrachters nicht auf sich. Der Betrachter braucht sich nicht umzusehen, um die sprechende Person zu finden. Dieses Ergebnis bestätigt sich mit den Ergebnissen von Magliano und Zacks (Magliano und Zacks, 2011), dass Diskontinuität in der Aktion größere Auswirkungen hat als Diskontinuität im Raum. Es entspricht auch den Regeln von Murch, dass die Geschichte wichtiger ist als der Raum (Murch, 1992).

4.3 Realisierung von Spaceline-Konstrukten

4.3.1 Guidelines für die Konstruktion einer Spaceline-Applikation

Aus der analysierten Literatur und den Ergebnissen der **Studien 7.3.1 bis 7.3.6** ergeben sich folgende Guidelines.

Regionen

Targetgrößen von weniger als 3° sollten vermieden werden. Bei der Nutzung von Eyetrackern muss bei der Auswahl der Targetgröße die Eyetracker-Genauigkeit berücksichtigt werden. Ist diese gering, sollte das Target größer sein.

Pointer und Trigger

Für die Pointing-Phase der Selektion sind kopf- und blickbasierte Techniken geeignet. Sie entsprechen dem natürlichen Selektionsvorgang in der realen Welt, in der zuerst ein Objekt erblickt wird, bevor damit agiert wird.

Um kopf- und blickbasierte Trigger zu realisieren, braucht es eine genaue Bestimmung der Parameter. Viele der Parameter variieren in ihrer Wirksamkeit zwischen den Personen. Die blink-Technik kann auf Dauer ermüdend sein, wenn sie häufig und lange verwendet wird.

Die beste Trigger-Technik in unserer Studie war die Nutzung einer *Controller-Taste*, auch wenn wir damit gerechnet hatten, dass der Controller in einer CVR-Anwendung als störend empfunden wird. Die Ergebnisse zeigen aber, dass die Nutzer daran gewöhnt sind und einfache Controller nutzen können, ohne sich ständig darüber bewusst zu sein.

Die Verwendung von Gestenerkennung ermöglicht die Nutzung der Hand zum Pointing, ohne ein zusätzliches Gerät zu halten. Es ist ebenfalls eine sehr natürliche Technik, mit dem Finger auf eine Rol zu zeigen. In mehreren Studien hat die Gestenerkennung jedoch noch nicht zuverlässig funktioniert (Atienza u. a., 2016; Pakkanen u. a., 2017). Durch die sich entwickelnde Technik werden solche Probleme in naher Zukunft gelöst sein.

Obwohl kopf- und blickbasierte Techniken für CVR-Erfahrungen geeigneter erscheinen, können Gesten zum Navigieren im Film verwendet werden (Rovelo u. a., 2014). Die Ergebnisse von Rovelo et al. führten zu einem benutzerdefinierten Gesten-Set für CVR-Movie-Player (Stop, Rückwärts, Vorwärts, Abspielen).

Eine weitere Möglichkeit für das Pointing ist die Verwendung von Sensordaten (Hassib u. a., 2017). Abhängig von der Verfassung des Benutzers (Emotion, Wachsamkeit) kann die nächste Szene ausgewählt werden.

Sensordaten können neben der Auswahl der nächsten Szene auch zur Aktivierung herangezogen werden. In einer meditativen Szene könnte beispielsweise die nächste Sequenz erst dann beginnen, wenn der Betrachter ein gewisses Entspannungslevel erreicht hat.

Transitions und Ausrichtung der Spaces

Wir fanden mehrere Transition-Methoden, die sowohl für dis-located als auch co-located Kamerapositionen verwendet werden können, wenn einige Richtlinien befolgt werden.

Die *toRol*-Methode kann verwendet werden, um den Benutzer zu wichtigen Rols zu führen. Wenn es eine Kontinuität in der Story gibt, bemerken die Zuschauer Diskontinuitäten in Raum und Richtung nicht. Bei Erkundungsszenen mit weniger hervorstechenden Rols oder ohne fortlaufende Story kann die *toRol*-Methode allerdings zu Desorientierung führen. Um dies zu vermeiden, sollte der Bereich, der nach der Transition gesehen wird, die Aufmerksamkeit des Betrachters auf sich ziehen. Die *toRol*-Methode funktioniert für auffällige Objekte auf Augenhöhe des Betrachters.

Die *wait*-Methode sollte für unauffälligere Elemente oder Bereiche, die weiter von der horizontalen Linie entfernt sind, bevorzugt werden. Diese Methode kann auch verwendet werden, wenn ein Space mehr als eine wichtige Rol enthält.

Die *keepFocus*-Methode sollte nicht mit *timeline-basierten* Transitions verwendet werden. Sie ist jedoch wichtig für den Wechsel zwischen Einstellungsgrößen. Dabei sollten *spaceline-basierte* Transitions verwendet werden. Es ist zu beachten, dass sich die Bildausschnitte vor und nach der Transition wesentlich voneinander unterscheiden. Bei *spaceline-basierten* Transitions ist die Blickrichtung vor und nach der Transition, so dass die Position der Kamera so bestimmt werden kann, dass die 30°-Regel eingehalten wird.

Vermeidung von Achsensprüngen

Erblickt der Zuschauer etwas auf der linken Seite, das sich vor der Transition auf der rechten Seite befand, hat dies ähnliche Auswirkungen wie das *Achsensprung-Problem* im traditionellen Film. In CVR kann diese Situation eintreten, auch wenn die Kamera die Aktionsachse nicht kreuzt. Um solche Transitions zu vermeiden, muss für *co-located* Aufnahmen die Blickrichtung vor und nach der Transition berücksichtigt werden. Dies ist für *spaceline-basierte* Transitions möglich. Bei *timeline-basierten* Transitions kann die Blickrichtung vor der Transition jedoch durch das Wissen, welche Bereiche Aufmerksamkeit erregen (z. B. Bewegungen oder Personen), abgeschätzt werden.

Zusammenfassend fanden wir die folgenden Guidelines, die nicht als starre Regeln, sondern als Empfehlungen zu verstehen sind. Sie basieren auf Experimenten mit ausgewählten Filmsequenzen, die nicht die Fülle der möglichen Filmstile abdecken. Sie repräsentieren das aktuelle Verhalten der Zuschauer, welches sich mit zunehmenden VR-Erfahrungen ändern kann.

Regionen: Targetgrößen von weniger als 3° sollten vermieden werden.

Pointing: Für die Selektion sind kopf- und blickbasierte Techniken geeignet.

Trigger:

- Um kopf- und blickbasierte Trigger zu realisieren, braucht es eine genaue Bestimmung der Parameter. Viele der Parameter variieren in ihrer Wirksamkeit zwischen den Personen
- Controller eignen sich gut als Trigger-Auslöser und haben in unseren Experimenten die Experience nicht beeinträchtigt.

Transitions:

toRol-Methode

- funktioniert für auffällige Objekte auf Augenhöhe des Betrachters,
- kann verwendet werden, um den Benutzer zu wichtigen Rols zu führen,
- kann auch für *co-located* Transitions verwendet werden.

keepFocus-Methode

- sollte nicht mit *timeline-basierten* Transitions verwendet werden, sondern mit *spaceline-basierte* Transitions,
- ermöglicht die Verwendung verschiedener Einstellungsgrößen,
- Bildausschnitte vor und nach der Transition sollten sich wesentlich voneinander unterscheiden.

wait-Methode

- sollte für unauffälligere Rols oder Bereiche, die weiter von der horizontalen Linie entfernt sind, bevorzugt werden,
- kann angewendet werden, wenn ein Space mehr als eine wichtige Rol enthält.

Vermeidung von Achsensprüngen: Für *co-located* Aufnahmen muss die Blickrichtung vor und nach der Transition berücksichtigt werden. Der Blick sollte nach der Transition auf eine story-relevante Rol gerichtet sein.

4.3.2 Design-Space für Spaceline Konstrukte

Aus der Analyse der Literatur und den Ergebnissen unserer Studien wurde ein Design-Space für das Spaceline-Konzept entwickelt (Tabelle 4-3). Jede der vier Dimensionen weist mehrere Subdimensionen auf, die anschließend zusammen mit möglichen Werten detailliert beschrieben werden. Ziel des Design-Spaces ist es, CVR-Entwickler bei der Kreierung interaktiver CVR-Applikationen zu unterstützen, in dem aufgezeigt wird, welche Dimensionen und Attribute im Designprozess berücksichtigt werden müssen.

Dimension		Attribute (Optionen)			
Region	Typ	out-Region	add-Region	in-Region	
	Position	Koordinaten			
	Form	klein	mittel	groß	
	Priorität	hoch	mittel	niedrig	
	Lifetime	permanent	begrenzt	sequenziell	
	Indikator	siehe Kapitel 2			
Pointer	Mount	Kopf	Blick	Controller	Hand
	Sichtbarkeit	subtil	apparent	unsichtbar	
	Form	Strahl	Kegel	3D-Objekt	
	Feedback	Pointer-Layout	Region-Layout	auditiv	haptisch
Trigger	Mount	Kopf	Blick	Controller	Hand
	Signal	Nicken/ Dwell	Dwell/ Blinzeln	Taste	Geste
	Feedback	visuell	auditiv	haptisch	keines
Transition	Ausrichtung	toRol	keepDir	keepFocus	personView
	Effekt	Blenden	Harter Schnitt	Überblendung	

Tabelle 4-3: Komponenten des Spaceline-Konzeptes. Die Tabelle zeigt für jede Komponenten die Subkomponenten und mögliche Attribute.

Typ und **Position** sind die Basiseigenschaften einer Region. Für das Spaceline-Konzept wurden in Abschnitt 4.2.1.2 *out*-, *add*- und *in*-Regionen eingeführt. Die Positionen dieser Regionen hängen vom Inhalt der CVR-Experience ab und bestimmen den Inhalt wesentlich mit. Sie werden mittels Koordinaten beschrieben.

Die Berücksichtigung der **Form** der Region ist entscheidend für den Grad der Schwierigkeit, eine Region zu adressieren. Obwohl *große* Targets leichter auswählbar sind, kann bei einigen Trigger-Techniken leicht ein Midas Touch passieren. *Kleine* Targets sind schwerer zu adressieren. Die Definitionen für *klein*, *mittel* und *groß* hängen von den verwendeten Techniken ab und sollten im Designprozess getestet werden.

Die **Priorität** einzelner Regionen kann unterschiedlich sein: Einige können notwendig für das Verständnis sein und der Wechsel zum nächsten Space sollte erst erfolgen, wenn die Region gesehen worden ist (*hohe* Priorität). Bei anderen Regionen kann es genügen, eine von mehreren gesehen zu haben (*mittlere* Priorität), weitere bieten zwar zusätzliche Informationen, sind aber für den Verlauf der Story nicht unbedingt notwendig (*niedrige* Priorität).

Regionen haben eine unterschiedliche **Lifetime**. Sie können *permanent* aktiv oder für ein *begrenzt*es Zeitintervall verfügbar sein. In einigen Fällen ist es auch denkbar, eine Reihenfolge zu definieren, in der der Betrachter die Bereiche gesehen haben muss (*sequenziell*).

Indikatoren sind nicht nur eine Charakteristik einer Region, sondern weisen selbst wieder verschiedene Unterdimensionen auf, die in Kapitel 2 ausführlich beschrieben wurden. Ein Indikator kann dem Zuschauer die Targetfindung erleichtern, aber in manchen Fällen das Seherlebnis beeinträchtigen. Es sind jedoch auch CVR-Applikationen ohne Indikatoren sinnvoll, in denen dem Benutzer nicht bewusst werden soll, dass ein Target aktiviert wird.

Pointer

Zum Zeigen auf eine Region (Pointing) ist ein Pointer erforderlich. Dieser Pointer kann z. B. mit der *Kopfrichtung*, dem *Blick*, einem *Controller* oder dem Zeigefinger (*Hand*) verbunden sein (**Mount**). In einer CVR-Applikation ist die Blickrichtung basierend auf Eye- oder Headtracking die natürlichste Art der Adressierung.

Je nach Anwendung kann der Pointer eine unterschiedliche **Sichtbarkeit** haben. Bei einigen Applikationen sollte er *subtil* oder sogar völlig *unsichtbar* sein, um das Seherlebnis nicht zu beeinträchtigen, bei anderen deutlich sichtbar (*apparent*), um die Selektion zu unterstützen.

In der Literatur findet man verschiedene Pointer-**Formen** für VR-Anwendungen: *Strahlen*, *Kegel* (Hand, 1997; Steed, 2006) und auch *3D-Objekte*, wie z. B. eine virtuelle Hand (Bowman u. a., 2005) oder ein Werkzeug. Ein *Strahl* ist präziser als ein *Kegel*, aber für kleine Objekte schwieriger zu verwenden. Für manche CVR-Applikationen sind auch *3D-Objekte* denkbar, wenn der Zuschauer eine aktive Rolle einnimmt und das *3D-Objekt* Teil der Geschichte ist.

Um den Zuschauern **Feedback** bei der Adressierung einer Region zu geben, kann der *Pointer* sein Aussehen (Farbe, Größe, Sichtbarkeit) in der *RoI* ändern. Alternativ ändert die *Region* selbst ihre Layout-Attribute (z. B. Farbe, Helligkeit, Umrandung), wenn der Pointer auf das Target trifft. Auch Geräusche (*auditiv*) oder Vibrationen (*haptisch*) sind mögliche Feedbacks.

Trigger

Sobald der Benutzer eine Region adressiert hat und eine Aktion getriggert werden soll, wird eine Technik mit diskreter Eingabe benötigt. Analog zum Pointing kann dieses Instrument der *Kopf*, der *Blick*, ein *Controller* oder eine *Geste* durch eine *Hand* sein (**Mount**). Auch andere Instrumente (z. B. *Sprache*) sind möglich.

Je nach genutztem Instrument sind unterschiedliche **Signale** als Auslösemechanismus anwendbar. Für den Kopf können *Nick*- oder andere *Kopfbewegungen* verwendet werden. Bei Anwendung der Dwell-Zeit wird das Target aktiviert, wenn die Kopfrichtung des Benutzers für ein bestimmtes Zeitintervall das Target adressiert. Dies ist auch für die Augen praktikabel. Eine andere blickbasierte Trigger-Technik ist z. B. das *Blinzeln*. Darüber hinaus kann das Ereignis per Handbewegung (*Geste*) oder sogar automatisch (sofort, ohne zusätzliche Aktionen) ausgelöst werden, sobald sich der Pointer auf dem Target befindet. Selbst *Sensordaten* (z. B. Puls) können zum Auslösen eines Ereignisses verwendet werden.

Es hängt von der Anwendung ab, ob das Triggern ein zusätzliches **Feedback** erfordert. Meist ist das Ergebnis des Triggerns schon ausreichend. Der Betrachter bemerkt, dass eine neue Szene startet oder ein Popup-Fenster geöffnet wird. Neben *visuellen* Feedbacks sind auch *akustische* (z. B. Piepton) und *haptische* (z. B. Vibration) Feedbacks anwendbar. In manchen Anwendungen ist dies in die CVR-Welt integrierbar, zum Beispiel das Geräusch einer knarrenden Tür.

Transition

Während einer Transition kann der Space in einer geeigneten Weise **ausgerichtet** werden. Je nach Ziel des Links sind unterschiedliche Richtungen möglich: *toRoI*, um auf eine *RoI* hinzuweisen, *keepFocus*, um ähnliche Effekte wie mit dem Wechsel von Einstellungsgrößen zu erzielen, *keepDir*, um die Richtung zu behalten und *personView*, um die Sichtweise eines Charakters anzunehmen.

Ähnlich wie in framed Filmen sind für den Space-Übergang verschiedene visuelle **Effekte** verwendbar, z. B. *Schwarzblenden* oder *Überblendung* für Diskontinuitäten in Zeit und Ort und *harte Schnitte* beim Wechsel von Einstellungsgrößen.

4.3.3 Konzept für einen CVR-Editor

Damit CVR-Applikationen auch ohne Programmierkenntnisse erstellt werden können, ist eine Unterstützung durch ein Editorprogramm erforderlich, welches ähnlich einem Filmschnittprogramm, die Komposition des Filmmaterials ermöglicht. Um Erfahrungen zu sammeln, welche Komponenten in einem CVR-Implementierungstool erforderlich sind und um Elemente des Spaceline-Konzeptes zu untersuchen, wurde der Prototyp eines CVR-Editors entwickelt. Die in diesem Prototyp implementierten Indikatoren wurden in Nutzerstudien untersucht und die gesammelten Erfahrungen genutzt, um ein Konzept für einen CVR-Editor zu entwickeln.

4.3.3.1 Prototyp

In Anlehnung an den vorgestellten Design-Space (Abschnitt 4.3.2) wurde der Prototyp eines CVR-Editors konzipiert und implementiert (Abbildung 4-10). Das primäre Ziel war dabei, Testanwendungen für Nutzerstudien flexibel gestalten zu können, um Eigenschaften von Regionen, Pointer und Indikatoren zu erforschen. Mit dem Editor können Spaces zu einem Spaceline-Konstrukt kombiniert werden. Dies geschieht durch die Definition von *out*- und *in*-Regionen und die Festlegung von *on*-/*off*-screen Indikatoren. Darüber hinaus können mit dem Tool weitere Interaktionen für Aktionsbereiche definiert werden, z. B. das Einbetten von Texten und Bildern.

Das Tool unterstützt eine Vorschau der erstellten VR-Applikation über HMD, ähnlich dem Wiedergabemodus eines herkömmlichen Schnittprogramms mit Timeline. Es können verschiedene Indikatortypen ausgewählt und in ihrem Erscheinungsbild an das Filmprojekt angepasst werden. Dadurch ist es möglich herauszufinden, wann Indikatoren benötigt werden und wie diese aussehen sollten, um das Filmerlebnis positiv zu beeinflussen.

Die implementierten Indikatoren sind Beispiele, wie Betrachter auf Regionen aufmerksam gemacht werden können. Es gibt jedoch mehr Möglichkeiten für *on*- und *off*-screen Indikatoren, die weitere Analysen benötigen. Die in den Studien untersuchten Indikator-Designs unterstützen die Interaktionen des Betrachters. Es sind jedoch auch subtile Regionen vorstellbar, in denen Space-Wechsel ohne Indikatoren ausgelöst werden.

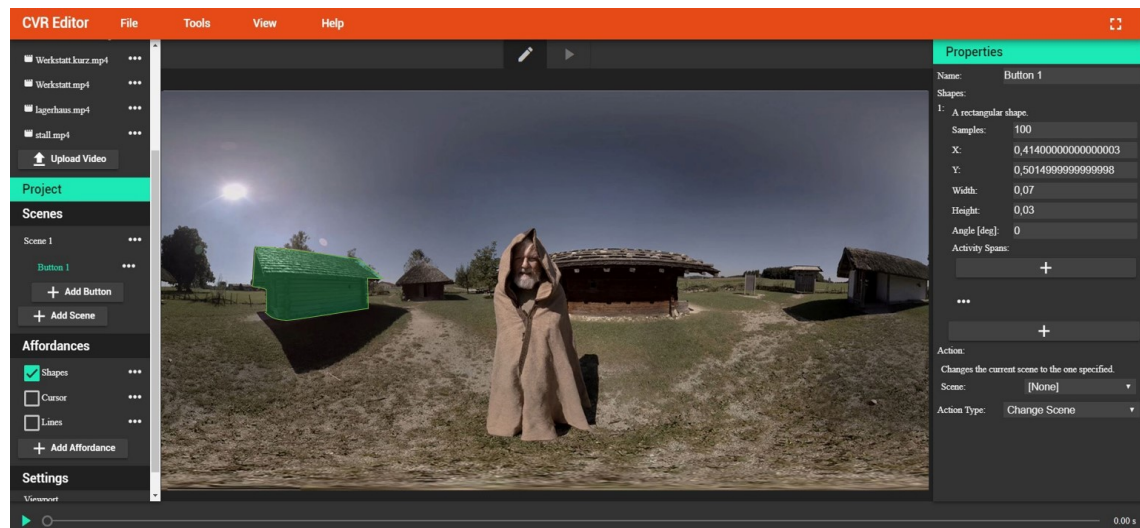


Abbildung 4-10: CVR-Editor: Tool zum Erstellen eines Spaceline-Konstrukts, dem Definieren der Regionen und Indikatoren. Links können die Elemente ausgewählt, rechts die Parameter definiert werden. Auf dem Filmbild ist ein Shape für eine der Hütten definiert (transparentes, grünes Overlay).

Der Prototyp des CVR-Editors wurde in den **Studien 7.3.3** und **7.3.4** genutzt, um verschiedene *on*- und *off*-screen Indikatoren zu untersuchen. Zur Visualisierung von *on*-screen Regionen wurden diese für eine Teilnehmergruppe mit Shapes versehen (Abbildung 4-11). Außerdem wurde getestet, ob eine

Visualisierung des Pointers durch einen Cursor und die Veränderung des Cursorlayouts ausreichend ist, um auf Rols aufmerksam zu machen. Bei dieser Methode ändert der Cursor seine Farbe und wird auffälliger, sobald er sich in einer *out*- oder *add*-Region befindet (Abbildung 4-12).



Abbildung 4-11: Beispiel für einen *on-screen* Indikator: Ein *Shape* hat die Form der RoI und ist sichtbar, sobald die RoI im Sichtfeld ist. Die Fläche ist teil-transparent, so dass der Filminhalt sichtbar bleibt (links). Sobald der Zuschauer in Richtung der RoI schaut, verringert sich die Transparenz und die Fläche ist auffälliger (rechts).



Abbildung 4-12: Beispiel für einen *on-screen* Indikator: Der *Cursor* zeigt die Mitte des Displays an und ist sehr subtil (links). In einer RoI wird die Farbe des Cursors auffälliger und der Cursor sichtbarer (rechts).

Der Vergleich von *Shape* (Abbildung 4-11) und *Cursor-Layout* (Abbildung 4-12) in **Studie 7.3.4** zeigte, dass durch die *Shape*-Anzeige die *Presence* zwar etwas geringer wird, aber mehr *add*-Regionen von den Teilnehmern wahrgenommen und aktiviert werden als mit der *Cursor*-Methode. Die *Shape*-Anzeige wurde von den Teilnehmern als verständlicher eingestuft.

Zur Untersuchung von *off-screen* Indikatoren wurden zwei Techniken implementiert: ein Kantenspot an der Seite des FoVs, welche der RoI näher ist und ein Strahl, der die Mitte des FoVs mit der RoI verbindet.



Abbildung 4-13: Beispiele für *off-screen* Indikatoren: Kantenspot (links), Strahl (rechts).

Die Verwendung von *off-screen* Indikatoren in **Studie 7.3.4** hatte keinen Einfluss auf die Anzahl der Interaktionen. Aus den Kommentaren der Teilnehmer geht hervor, dass die *off-screen* Indikatoren (Abbildung 4-13) für die meisten Teilnehmer nicht selbsterklärend waren. Es bedarf weiterer Untersuchungen wie *off-screen* Indikatoren für CVR gestaltet sein sollten.

Der Prototyp diene nicht nur als Tool für Studien zum Spaceline-Konzept, sondern darüber hinaus zum Sammeln von Erfahrungen, wie ein Programm zu Erstellung spaceline-basierter CVR-Applikationen aussehen sollte. Aufgabe eines solchen CVR-Editors ist es, einen CVR-Entwickler bei der Arbeit an einem Spaceline-Konstrukt zu unterstützen.

4.3.3.2 Komponenten

Aus dem Design-Space in 4.3.2 (Tabelle 4-3) ergeben sich die folgenden Hauptkomponenten eines CVR-Editors: *Space*, *Pointer* und *Trigger* (Abbildung 4-14). Die Design-Space-Dimensionen *Region* und *Transition* sind beide in der *Space*-Komponente enthalten. Jede Komponente weist mehrere Attribute auf, die zum Großteil den Optionen im Design-Space entsprechen und in Abschnitt 4.3.2 beschrieben wurden.

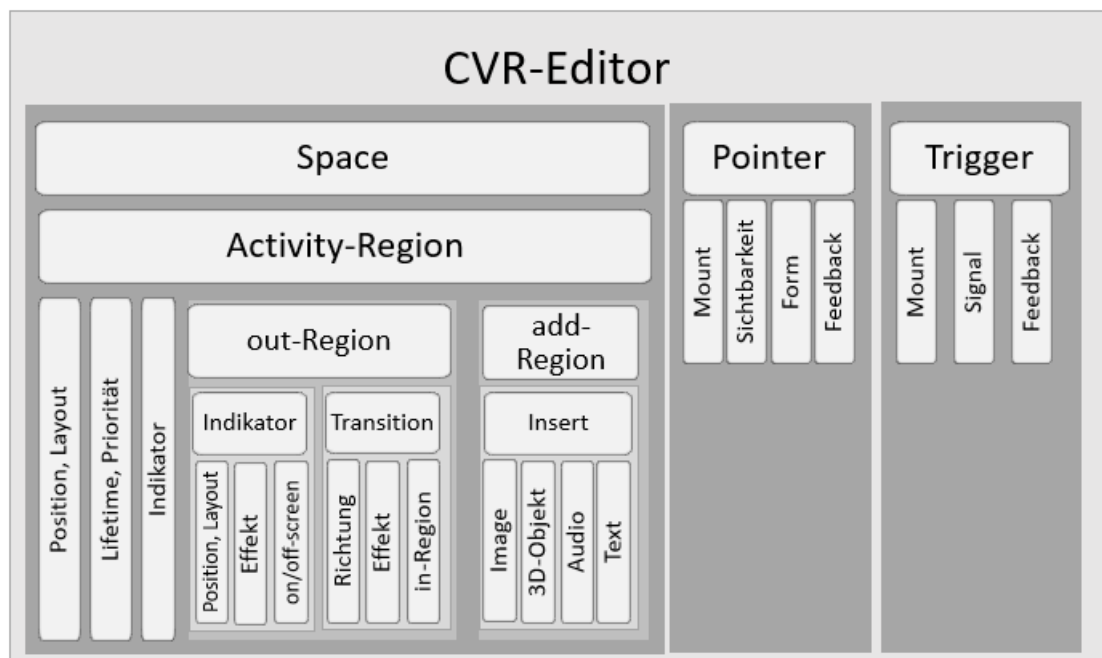


Abbildung 4-14: Komponenten des CVR-Editors.

In Anlehnung an die Syntax von UML-Klassendiagrammen werden in Abbildung 4-15 die einzelnen Komponenten dargestellt und zueinander in Beziehung gesetzt.

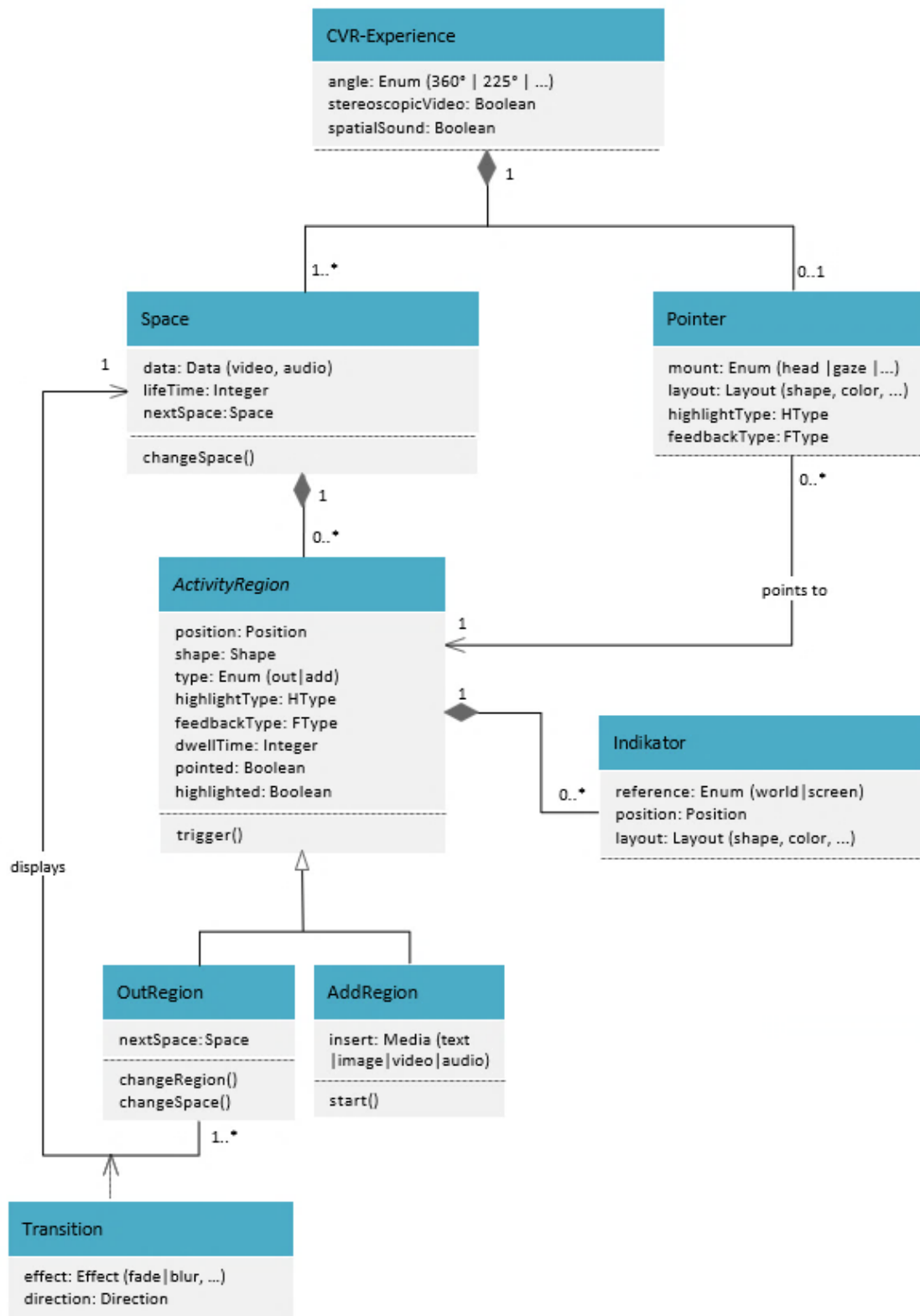


Abbildung 4-15: Beziehungen der Komponenten und Unterkomponenten zueinander.

4.3.3.3 Design-Überlegungen

Die Auswahl der Attribute für die jeweilige CVR-Applikation ist ein kreativer Prozess, der Kenntnisse und Erfahrungen sowohl im Film, wie auch im Design von VR-Erfahrungen erfordert. Dabei stehen folgende Aufgaben im Vordergrund:

- *out-* und *add*-Regionen: Welche Regionen werden benötigt (Position, Form)? Soll durch Indikatoren auf die Regionen verwiesen werden, wenn ja wie (Indikatoren)? Wann sollten die Regionen aktivierbar sein (Lifetime) und ist ihre Aktivierung notwendig (Priorität)?
- Transition: Wie erfolgt der Übergang zwischen den Spaces? Wohin schaut der Zuschauer danach (Richtung)? Ist ein Effekt (z. B. Blende) notwendig?
- Inserts: Welche Zusatzinformationen werden vermittelt? Sollten diese in Filmen, Bildern, Texten oder Audios bereitgestellt werden?
- Pointer: Wie soll eine Region ausgewählt werden (z. B. blickbasiert)?
- Trigger: Wie soll ein Szenenwechsel oder eine Einbettung ausgelöst werden?

Die fertige Anwendung kann dem Zuschauer als Applikation für unterschiedliche VR-Geräte zur Verfügung gestellt werden. Aber auch ein spezielles Dateiformat für CVR-Anwendungen, unter anderem einem CVR-Player, ist denkbar.

4.3.3.4 CVR-Containerformat

Um die für eine CVR-Applikation notwendigen Daten zu speichern und zwischen Programmen austauschen zu können, wird ein CVR-Containerformat benötigt, welches alle notwendigen Daten enthält, die für die CVR-Anwendung notwendig sind: die Bild- und Audiodaten jedes Spaces, die Beschreibung der Regionen, des Pointers, der Trigger-Ereignisse und Transitions.

Containerformaten bieten die Möglichkeit, verschiedene Dateien unterschiedlicher Formate gemeinsam mit Metadaten zu speichern.

Video-Containerformate wie Quicktime (.mov), AVI (.avi) und MPEG-4/Part 14 (.mp4) enthalten Audio- und Videostreamdateien, die so miteinander verbunden sind, dass die Datei mit einer entsprechenden Player-Applikation als ganzer Film abspielbar ist, wenn die für die Streamdateien benutzten Codecs vorhanden sind (Mücher, 2016, 2014a, 2013). Die Video- und Audio-Dateien enthalten nur noch das tatsächlich im Film gezeigte Material.

Diese Formate werden momentan für einen Großteil der produzierten omnidirektionalen Filme benutzt. Sie lassen sich mit herkömmlichen Filmschnittprogrammen produzieren und Schnittprogramme wie Adobe Premiere und DaVinci Resolve bieten über Plug-Ins zusätzliche Möglichkeiten der Bearbeitung. Allerdings sind die produzierten Filme immer linear und timeline-basiert. Um das CVR-Konzept umsetzen zu können, braucht es ein Containerformat, in dem die Videostreams in voller Länge enthalten sind und die Wechsel in Abhängigkeit vom Blickverhalten ausgelöst werden. Dafür eignen sich Containerformate, die im Allgemeinen dem Produktionsprozess vorbehalten sind.

Austauschformate wie das Material Exchange Format (MXF), das Open Media Framework (OMF) und das Advanced Authoring Format (AAF) enthalten ebenfalls Video-, Audio und Metadaten und sind für den Austausch während der Bearbeitungsphase wichtig. Sie werden von zahlreichen professionellen Audio- und Videobearbeitungsprogrammen unterstützt. Die Video- und Audiodateien sind dabei meist in ungekürzter Form im Container enthalten, so dass Positionen von Schnitten jederzeit verändert werden können. MXF enthält keine Informationen über Bildeffekte und wird deshalb meist für den Austausch von Rohschnitten oder schon gerenderten Beiträgen verwendet (Mücher, 2014b). In AAF-Containern können auch Effekte abgespeichert werden, allerdings enthalten diese auch alle Bearbeitungsschritte (Mücher, 2018). OMF-Container können in begrenztem Umfang Informationen zu Schnitten enthalten und nach einer Konsolidierung das schnittrelevante Material (Mücher, 2012).

Auch wenn Austauschformate ursprünglich für den Produktionsprozess entwickelt wurden, ermöglichen sie für CVR die Speicherung von Daten, die in Abhängigkeit von der Blickrichtung von einem Player abgespielt werden können. Abbildung 4-16 stellt die Bestandteile eines CVR-Containers grafisch dar.

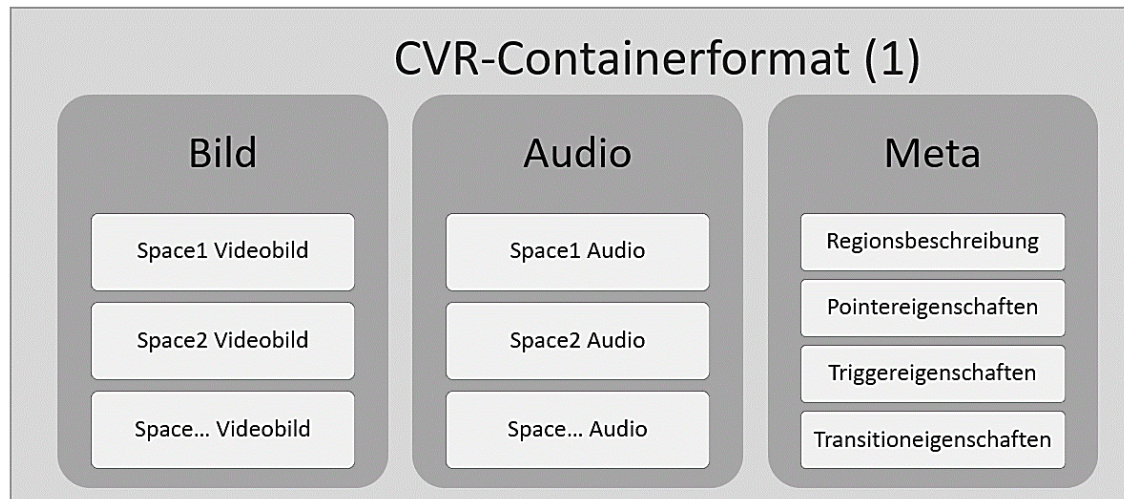


Abbildung 4-16: Container-Dateiformat für CVR-Spaceline-Anwendungen. Darin enthalten sind die Bild- und Audio-Daten jedes Spaces, die Definitionen der Regionen und Transitions, sowie auch die Eigenschaften des Pointers und der Triggerereignisse.

In Abschnitt 5.4 werden basierend auf den Ergebnissen des Kapitels 5 weitere Komponenten zu diesem Format hinzugefügt und ein Konzept für einen CVR-Player vorgestellt.

4.4 Fazit und Ausblick

In diesem Kapitel wurde das von uns entwickelte Konzept einer Spaceline in Analogie zur traditionellen Timeline vorgestellt. Filmbegriffe wie Einstellung und Sequenz wurden erweitert und im neuen Kontext erläutert. Neue Begriffe wie *Spaces*, *Spaceline*, *in-*, *out-* und *add-*Regionen wurden eingeführt und *on-screen* und *off-screen* Indikatoren definiert.

Es wurden relevante Erkenntnisse anderer Bereiche beschrieben und auf ihre Übertragbarkeit in CVR überprüft. Unser Spaceline-Konzept unterstützt Filmemacher, CVR-Applikationen mit dynamischen Handlungssträngen zu erstellen, in denen Szenenwechsel von den vom Betrachter gesehenen Regionen abhängen.

Erste Tests (**Studien 7.3.1 bis 7.3.6**) zeigten, dass das Spaceline-Konzept geeignet ist, nichtlineares Storytelling in CVR umzusetzen. Es bedarf jedoch weiterer Nachforschungen, um herauszufinden, welche Art von Interaktionsmethoden das Benutzererlebnis unterstützen, ohne den Erzählfluss zu stören.

CVR ist ein relativ neues Medium, das neue Techniken und eine neue Filmsprache benötigt, die die zusätzlichen Freiheiten der Betrachter berücksichtigt, die Blickrichtung und damit den Bildausschnitt selbst zu ändern. Um Erkenntnisse über das Verhalten der Zuschauer in CVR zu gewinnen, haben wir verschiedene Techniken zum Ausrichten von Spaces untersucht. Diese Methoden können zur Aufmerksamkeitslenkung, zum interaktiven Geschichtenerzählen und als Stilmittel in CVR verwendet werden. Es wurde erläutert, für welches Szenario, welche Methode angewendet werden kann.

Um interaktive Handlungsstränge in CVR zu realisieren, ist sowohl die Raum- als auch die Zeitkomponente erforderlich. Bei Verwendung des Spaceline-Konzepts sollten Filmemacher ebenfalls in der Lage sein, ein Zeitlimit für eine Aufnahme festzulegen.

Auch wenn die Ergebnisse nur das aktuelle Verhalten der Zuschauer, die oft wenig VR-erfahren sind, widerspiegeln können, ist es wichtig zu wissen, wie sich die Zuschauer in verschiedenen Szenarien fühlen, um eine Filmsprache für die filmische virtuelle Realität zu entwickeln.

5

How to Watch: Nutzererlebnis in CVR

„Bei der ständigen Entwicklung, in der sich die Dinge gegenwärtig noch befinden, wo der Streit „für“ und „wider“ so heftig tobt, wo täglich neue Gebiete vom Kino erobert werden, wo ihm täglich neue Entwicklungsmöglichkeiten zugesprochen und ebenso oft jede Berechtigung und Eignung dazu abgestritten wird, wo selbst die Mehrzahl derjenigen, denen die kinematographische Technik ein selbstverständlicher Zubehör zum Lebensapparat geworden ist, nicht einmal wissen, welche Stellung sie der Erscheinung im ganzen gegenüber einnehmen, da ist es natürlich nur möglich, ein Abbild der Lage zu geben und die prinzipielle Beurteilung einer späteren Zeit zu überlassen.“

Emilie Altenloh (Altenloh, 1913)

Beim Entwickeln von Konzepten für nutzerfreundliche CVR-Applikationen steht man vor mehreren Herausforderungen. Das Anschauen von Filmen ist für viele Menschen ein Gemeinschaftserlebnis, welches in CVR schwer realisierbar ist, da sich die Zuschauer durch das HMD voneinander isoliert fühlen. Es ist notwendig zu untersuchen, ob und wie die Nutzung von HMDs für gemeinsame CVR-Erlebnisse umgesetzt werden kann. Ein zweiter wesentlicher Aspekt ist die Gestaltung der CVR-Anwendung in einer Weise, dass sie auch ohne intensive Drehbewegung erfahrbar ist und ein entspanntes Filmerlebnis ermöglicht. Und schließlich müssen diese Filmerlebnisse barrierefrei gestaltet werden und damit auch Menschen zugänglich sein, die Zusatzdienste benötigen. Für alle drei Aspekte werden im Folgenden Herausforderungen identifiziert und Ideen vorgestellt. Es werden erste Lösungsansätze untersucht, die Ergebnisse diskutiert und in Designideen für einen CVR-Player zusammengefasst.

Dieses Kapitel enthält bereits veröffentlichtes Material aus folgenden Publikationen:

- Rothe, S., Montagud, M., Mai, C., Buschek, D., Hußmann, H., 2018. Social Viewing in Cinematic Virtual Reality: Challenges and Opportunities, in: International Conference on Interactive Digital Storytelling - ICIDS 2018.
- Rothe, S., Tran, K., Hußmann, H., 2018. Dynamic Subtitles in Cinematic Virtual Reality. in: Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video - TVX
- Rothe, S., Tran, K., Hussmann, H, 2018. Positioning of Subtitles in Cinematic Virtual Reality. in: International Conference on Artificial Reality and Telexistance & Eurographics Symposium on Virtual Environments - ICAT-EGVE

Außerdem beinhaltet Abschnitt 5.1 Material aus einer eingereichten, aber noch nicht veröffentlichten Arbeit über CVR als soziale Erfahrung, die ich als Hauptautor gemeinsam mit Alexander Schmidt, Mario Montagud, Daniel Buschek und Heinrich Hußmann in englischer Sprache verfasst habe.

Die Verwendung eines HMDs ermöglicht es dem Betrachter, einen sphärischen Film besonders immersiv wahrzunehmen. Bei der Präsentation von CVR-Experiences müssen allerdings mehrere Aspekte beachtet werden: Zum einen erschwert die Isolation des Zuschauers durch das HMD den Einsatz dieses Mediums als soziales Erlebnis. Zum anderen ist das Anschauen längerer CVR-Anwendungen oft anstrengend, was nicht nur auf die derzeit verfügbaren Geräte zurückzuführen ist. Nicht zuletzt sind Überlegungen notwendig, um die erforderliche Barrierefreiheit des Mediums zu realisieren.

In *Social Viewing Szenarien* schauen mehrere Nutzer (*Co-Watcher*) gleichzeitig einen Film und haben dabei die Möglichkeit miteinander zu kommunizieren. Der Film kann auf unterschiedlichen Geräten betrachtet werden. Die Kommunikation kann direkt erfolgen, aber auch über soziale Netzwerke, Messaging Dienste oder spezielle Plattformen. In Abschnitt 5.1 werden Herausforderungen und Designaspekte identifiziert, die für die Unterstützung des sozialen Zusammengehörigkeitsgefühls (engl.: *togetherness*) beim gemeinsamen Betrachten von Filmen von Bedeutung sind. Es wird ein Überblick über den aktuellen Forschungsstand gegeben, und Herausforderungen werden identifiziert. Für diese Herausforderungen werden mögliche Ansätze vorgeschlagen, von denen einige in Nutzerstudien untersucht wurden.

Um möglichst viel vom angebotenen Bildmaterial in CVR anschauen zu können, muss der Zuschauer seinen Kopf ständig in unterschiedliche Richtung drehen. Das ist auf Dauer anstrengend und erzeugt zusätzlich das Gefühl - trotz des ständigen Umherschauens - etwas zu verpassen. In Abschnitt 5.2 werden deshalb verschiedene Methoden analysiert, die das Umherschauen *komfortabler* machen. Außerdem wird untersucht, ob der gesamte 360°-Raum für eine immersive CVR-Experience notwendig ist, und wie sich kleinere Bereiche (180° und 225°) auf das Zuschauererlebnis auswirken.

Auch ein *barrierefreier* Medienzugang ist in CVR erschwert. Dieses Thema wird in Abschnitt 5.3 in einem ersten Ansatz aufgegriffen. Da diese Problematik sehr umfangreich ist, können dabei nur Teilaspekte näher betrachtet werden. Die Themen Untertitel und Gebärdendolmetscher wurden anhand von Nutzerstudien näher untersucht. Für andere Aspekte können wir erste Ansätze anbieten, die sich aus den Ergebnissen und Erfahrungen anderer Studien ableiten lassen.

Jeder der Abschnitte 5.1 - 5.3 ist gleich aufgebaut: Nach einer Darlegung der theoretischen Hintergründe, in der auf themenrelevante Publikationen aus Psychologie, HCI und Medienwissenschaften eingegangen wird, werden die Herausforderungen für die jeweilige Problematik erläutert. Die von uns implementierten Methoden, die Lösungen für die beschriebenen Herausforderungen bieten, werden vorgestellt und analysiert. Dabei werden die Ergebnisse der Nutzerstudien aus Kapitel 7 zusammengefasst und die Vorteile und Einsatzszenarien der Methoden diskutiert. Einzelne Abschnitte enthalten zusätzlich Vorschläge für weitere Forschungsaspekte.

Das Kapitel schließt mit einem Konzept für CVR-Player Applikationen, in dem die Erkenntnisse der vorangegangenen Abschnitte zusammengefasst werden.

5.1 Gemeinsame CVR-Anwendungen

Im Gegensatz zum herkömmlichen Kino oder Fernsehen verfügt jeder CVR-Zuschauer über ein eigenes Display und ist beim Anschauen des Films über HMD von der Umgebung isoliert. Das gemeinsame Filmerlebnis stellt deshalb besondere Anforderungen.

5.1.1 Theoretischer Hintergrund

Social Viewing

In verschiedenen Arbeiten wurde die Relevanz von Social Viewing Szenarien diskutiert (Harboe u. a., 2008b; Nathan u. a., 2008; Shin und Kim, 2015; Voorveld und Viswanathan, 2015) und soziale Aspekte in kollaborativen Videoszenarien untersucht (Kim u. a., 2018; Shin und Kim, 2015). Einige Ansätze thematisierten die Kommunikation zwischen Menschen, die sich an unterschiedlichen Orten aufhalten und gemeinsam fernsehen: *2BeOn* (Abreu u. a., 2002) erweiterte TV-Sendungen mit Online-Kommunikationsdiensten wie Instant Messaging und Videokonferenzen. *Amigo TV* (Godon u. a., 2004) ergänzte das gemeinsame Fernsehen an verschiedenen Orten durch Voice-Chat, Text-Chat und einzelne Emoticons. *SocialTV* (Harboe u. a., 2008a) und *SocialTV 2* (Harboe u. a., 2008b) weisen die Zuschauer darauf hin, welche Fernsehsendung von den anderen Gruppenmitgliedern gesehen werden, und ermöglichen das Senden von Nachrichten.

Eine Reihe von Arbeiten setzt sich mit den sozialen Aspekten solcher Szenarien auseinander: Weisz et al. (Weisz u. a., 2007) integrierten Text-Chat in eine Social-Viewing-Anwendung und stellten einen positiven Einfluss auf die sozialen Beziehungen zwischen den Zuschauern fest. Allerdings fühlten sich diese vom Videoinhalt abgelenkt, während sie miteinander kommunizierten. Das Hinzufügen natürlicher Pausen im Video konnte das Gefühl der Ablenkung verringern. Voice-Chats erzeugen ein größeres Zusammengehörigkeitsgefühl als Schrift-Chats (Geerts u. a., 2011). Sprache kann außer Texten, auch die Richtung des Sprechers, Stimmungen und Gefühle übertragen. Obwohl Voice-Chat das Zusammengehörigkeitsgefühl erhöht, kann das Seherlebnis aufgrund von Ablenkung verringert sein (Geerts u. a., 2011).

Beim gemeinsamen Anschauen sphärischer Videos auf Tablets in co-located Umgebungen beobachten die Teilnehmer die körperlichen Bewegungen der anderen, um daraus deren Blickrichtung abzuleiten (Tang und Fakourfar, 2017). Diese Strategie ist beim Tragen von HMDs nicht realisierbar.

Virtuelle Zusammengehörigkeit, soziale Presence

Während mit *räumlicher Presence* das „Gefühl, dort zu sein“ bezeichnet wird, beschreibt *soziale Presence* das „Gefühl, zusammen zu sein“ (De Greef und IJsselsteijn, 2001). In der Literatur werden verschiedene Definitionen beider Begriffe und auch unterschiedliche Messmethoden verwendet (Skarbez u. a., 2017). Der Begriff Presence, ohne weitere Attribute, bezieht sich im Allgemeinen auf die räumliche Presence. Soziale Presence hängt von kommunikativen Signalen (IJsselsteijn u. a., 2000), von Interaktionen (Schultze und Brooks, 2019) sowie vom Aufgabentyp und weiteren Aspekten ab (Oh u. a., 2018).

Im Gegensatz zur räumlichen Presence, setzt das Gefühl des Zusammenseins in einer virtuellen Welt (soziale Presence, virtuelle Zusammengehörigkeit) die Anwesenheit anderer Personen voraus. Die virtuelle Zusammengehörigkeit wird sowohl von dem Gefühl, in der virtuellen Welt zu sein, beeinflusst als auch von der Kommunikation zwischen den Nutzern in der virtuellen Welt (Durlach und Slater, 2000). Oh et al. (Oh u. a., 2018) geben einen Überblick über Definitionen und Konzepte sozialer Presence. Basierend auf 152 Studien werden verschiedene Faktoren kategorisiert und diskutiert. Die Ergebnisse zeigen, dass Tiefencues (Hinweise zur Wahrnehmung räumlicher Tiefe, engl. depth cues), Audioqualität, haptisches Feedback und Interaktivität die soziale Presence erhöhen können. Es wird betont, dass soziale Presence nicht nur von der Umgebung abhängt, sondern auch von den Personen, die an dem Prozess beteiligt sind. Haptische Kommunikation in gemeinsam genutzten virtuellen Umgebungen kann das Zusammengehörigkeitsgefühl verbessern (Ho u. a., 1998).

Kollaboration in virtuellen Umgebungen

Bereits in den 1990er-Jahren wurde das Thema der Kollaboration in VR diskutiert (Carlsson und Hagsand, 1993; Margery u. a., 1999; Normand u. a., 1999). Projekte wie DIVE (Distributed Interactive Virtual Environment) (Carlsson und Hagsand, 1993) und COVEN (COLlaborative Virtual ENvironments) (Normand u. a., 1999) legten den Grundstein für die heutige Forschung zu diesem Thema.

CollaVR, ein Tool für Filmemacher, welches gemeinsames Arbeiten an CVR Projekten über HMDs ermöglicht, wurde von Nguyen et al. (Nguyen u. a., 2017) eingeführt. Es ist für die Kommunikation von Personen implementiert, die kollaborativ arbeiten, und nicht als Zuschauererlebnis. Maus und Tastatur werden als Eingabegeräte genutzt und eine grafische Oberfläche zeigt Timeline und Schaltflächen für die nutzbaren Funktionen an. Für die Interaktion werden Voice-Chat und Visualisierungen der Sichtfelder genutzt. Das FoV des Co-Watchers ist umrahmt und sichtbar, wenn sich die beiden FoVs überlappen. Andernfalls wird ein Pfeil angezeigt, der in die Richtung des FoVs des Co-Watchers weist. Ein weiteres Beispiel für kollaboratives Arbeiten sind VR-Videokonferenzen (Gunkel u. a., 2018, 2017).

Beim kollaborativen Inspizieren von 3D-Netzwerkdaten mittels CAVE und HMD wurden keine wesentlichen Unterschiede bezüglich der verbalen Kommunikation und dem gemeinsamen Fokus zwischen diesen beiden Umgebungen festgestellt (Cordeil u. a., 2017). Dabei wurden Leap-Motion-Sensoren verwendet, um das Zeigen auf Rols auch in der HMD-Umgebung zu ermöglichen, indem die Positionen der Finger visualisiert wurden. Zusätzlich wurden die FoVs der Benutzer angezeigt. In den Experimenten lösten Benutzer Aufgaben mit dem HMD schneller als in der CAVE.

Dorta et al. (Dorta u. a., 2016) verglichen kollaboratives Filmschauen in einem CAVE-ähnlichen begehbaren System (Hyve-3D) mit einem HMD-Szenario. Die Ergebnisse der Studie zeigten, dass HMDs ein höheres Gefühl der Presence hervorrufen, die Kommunikation zwischen den Zuschauern jedoch erschweren. Ein Grund dafür ist die Schwierigkeit zu wissen, wohin die andere Person schaut. Selbst wenn begehbare Systeme für soziale CVR-Erlebnisse gut geeignet erscheinen, sind sie selten verfügbar und nur für öffentliche Räume anwendbar. In unserer Arbeit suchten wir nach Lösungen für Szenarien im privaten Raum und konzentrierten uns auf die Verwendung von HMDs.

Videoplayer für sphärische Filme

Videoplayer zum Anschauen sphärischer Videos auf Monitoren von Desktop-PCs wurden von Chambel et al. (Chambel u. a., 2011) und Neng und Chambel (Neng und Chambel, 2010) betrachtet. Sie führten Techniken ein, um die Position des eigenen FoVs zur besseren Orientierung im gesamten 360°-Bild anzuzeigen. Bei Verwendung eines HMDs unterstützt das eigene Körpergefühl für die durchgeführten Bewegungen die Orientierung, so dass ein zusätzliches Fenster wie bei Desktop-Anwendungen nicht erforderlich ist. Dennoch könnte diese Methode übernommen werden, um das FoV des Co-Watchers anzuzeigen.

Montagud et al. (Montagud u. a., 2018) entwickelten einen barrierefreien 360°-Player für sphärische Filme mit einer personalisierten Darstellung der Zugriffsdienste (Untertitel, Audiodeskription und Gebärdendolmetscher). Dafür wurden verschiedene Methoden der Aufmerksamkeitslenkung implementiert, die den Benutzern helfen, sprechende Personen oder Aktionen zu finden.

Synchronisation

Die Synchronisierung der Medienwiedergabe auf den beteiligten Geräten ist eine wichtige Voraussetzung für kollaborative Filmerlebnisse (Boronat u. a., 2018; Montagud u. a., 2012). Dies umfasst die Entwicklung und Anpassung der entsprechenden Protokolle und Überwachungsalgorithmen. Die Mediensynchronisation muss auch nach der Anwendung von Steuerbefehlen (z. B. Play, Stop) erhalten bleiben. Wersync (Belda u. a., 2015; Montagud u. a., 2015) ist eine webbasierte Plattform für verteilten Medienkonsum und bietet neben der Synchronisation auch soziale Interaktion zwischen remote Nutzern. Für die Kommunikation wurde ein integrierter Text-Chat implementiert. Da das Eingeben von geschriebenen Texten in VR-Geräten schwierig ist, werden in dieser Arbeit andere Kommunikationstechniken untersucht.

5.1.2 Herausforderungen für Social Viewing in CVR

Im Folgenden werden sieben zentrale Herausforderungen beschrieben, wir für wichtig halten. Sie basieren auf unseren Studien-Erfahrungen, Diskussionen mit CVR-Entwicklern und den Ergebnissen der studierten Literatur.

Herausforderung 1 – Kommunikation: Ein zentrales Thema in Social Viewing Szenarien traditioneller Filme ist der Kommunikationskanal, der für die Interaktion verwendet wird. Dies gilt auch für CVR. Selbst wenn co-located Zuschauer ohne zusätzliche technische Hilfsmittel sprachlich kommunizieren können, braucht die sprachliche Kommunikation nähere Untersuchungen und möglicherweise Ergänzungen, da es z. B. schwierig ist, auf Details hinzuweisen oder Gefühlsregungen des Co-Watchers wahrzunehmen.

Herausforderung 2 - Kenntnis des Viewports: Eines der Hauptprobleme beim gemeinsamen Betrachten von CVR-Filmen über HMDs ist das unterschiedliche FoV der Benutzer und die fehlende Kenntnis des FoVs des anderen. Worauf bezieht sich der Kommentar eines Nutzers? Warum lacht der andere? Wie kann ein Betrachter auf Details im Film hinweisen, die nicht notwendigerweise im FoV des anderen sind? Nicht zu wissen, wohin der Co-Watcher in der sphärischen Szene schaut, kann eine Kommunikationsbarriere darstellen und zu Verständigungsschwierigkeiten und Missverständnissen führen.

Herausforderung 3 – Soziale Presence: Für Social Viewing ist es wichtig, das Gefühl des „Zusammenseins“ hervorzurufen – d. h. den Film nicht allein zu sehen. Beim gemeinsamen Anschauen eines Films im Kino oder Fernsehen wird der Co-Watcher am Rande des Blickfeldes wahrgenommen. Obwohl „stille“ Gefühle (z. B. Traurigkeit) vom anderen Benutzer nicht gehört werden, können sie durch Körperhaltungen oder Gesten erkannt oder abgeleitet werden. Dies ist in CVR nicht möglich.

Herausforderung 4 – Barrierefreiheit: Das gemeinsame Anschauen von CVR-Filmen sollte auch für Personen möglich sein, die zusätzliche Dienste benötigen, z. B. Untertitel / Gebärdensprache für Gehörlose und Hörbehinderte oder Audiobeschreibung für Blinde und Sehbehinderte. Auch benötigen körperbehinderte, die nicht ohne Weiteres ihren Kopf in alle Richtungen drehen können, zusätzliche Techniken. Eine Social Viewing CVR-Applikation sollte diese Möglichkeiten bieten.

Herausforderung 5 – Interaktion: Um miteinander zu kommunizieren und sich gegenseitig wahrzunehmen, sind Interaktionsmethoden erforderlich. Dabei muss beachtet werden, dass die Eingabetechniken das Seherlebnis nicht beeinträchtigen. Beispielsweise können grafische Menüs oder Tastaturen in der virtuellen Welt die Presence und Immersion verringern und das Seherlebnis mindern.

Herausforderung 6 – Synchronisation: Für das gemeinsame Anschauen von Filmen müssen alle beteiligten Benutzer die gleichen Ereignisse zur gleichen Zeit wahrnehmen können, damit inkonsistente Interaktionen und frustrierende Situationen vermieden werden (z. B. Jubel eines Co-Watchers über ein erzieltes Tor, bevor die Torsequenz auf dem anderen Display angezeigt wird).

Herausforderung 7 – Multiuser-Umgebungen: Auch wenn für erste Ansätze Szenarien untersucht werden, in denen sich zwei Personen gemeinsam einen sphärischen Film über HMD ansehen, sind auch Szenarien mit mehr Personen relevant, z. B. Schulkassen oder Präsentationen in Museen. Denkbar ist auch eine kinoähnliche Situation. In einer solchen Konstellation können visuelle und akustische Informationen aller Benutzer zu einer Überladung führen. Andere Szenarien sind asymmetrische Umgebungen: Die Teilnehmer verwenden verschiedene Arten von VR-Geräten oder sogar einen zusätzlichen Desktop. Auch die Heterogenität von Geräteumgebungen muss bei der Entwicklung interaktiver und sozialer CVR-Experiences berücksichtigt werden.

5.1.3 Methoden für kollaborative CVR-Erlebnisse

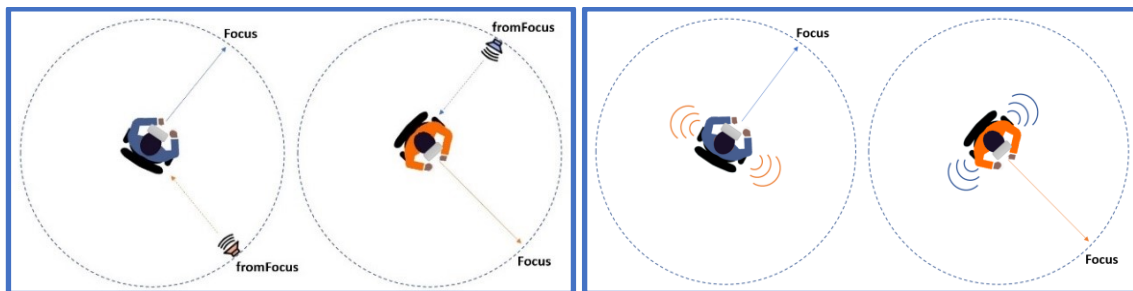
Jede der identifizierten Herausforderungen erfordert geeignete Designrichtlinien und Erkenntnisse für die Entwicklung von CVR-Applikationen. In diesem Kapitel werden verschiedene Ansätze und Methoden vorgestellt und miteinander verglichen.

5.1.3.1 Nutzerstudie: Voice-Chat: *spatial*-Methode, *nonspatial*-Methode

Sprache kann Texte, Stimmungen, Gefühle und auch die Richtung des Sprechers übertragen. *Voice-Chat* ist eine Möglichkeit, in remote Umgebungen zu kommunizieren. Um dies in CVR umzusetzen, gibt es verschiedene Optionen: Die Stimmen können räumlich in der virtuellen Welt positioniert werden, aber auch fest mit dem HMD verbunden sein – ähnlich wie world- und screen-referenzierte visuelle Elemente. Im Falle eines räumlichen Voice-Chats, ist es wichtig, zu entscheiden, aus welcher Richtung dieser kommen soll. Eine Variante wäre, die Quelle des Sounds auf eine virtuelle Position neben dem Zuschauer (der virtuellen Kamera) zu legen. Dies entspricht der Situation des Nebeneinandersitzens. Eine andere Möglichkeit ist, die Soundquelle auf dem Zentrum des Viewports der sprechenden Person zu positionieren. Die Stimme kommt dann aus der Richtung, in die der Sprecher blickt (Abbildung 5-1 a). Der Vorteil dieser Methode besteht darin, dass Personen es gewohnt sind, in die Richtung zu schauen, aus der ein Geräusch kommt. Auch wenn die erste Variante natürlicher erscheint, stellten wir in einer informellen Vorstudie fest, dass die Zuschauer die zweite Variante bevorzugten.

Da die Chat-Stimmen nicht zum Filmerlebnis gehören, aber wie ein Element der virtuellen Welt wirken, ist es möglich, dass der räumliche Klang die Teilnehmer irritiert. Aus diesem Grund wurden in **Studie 7.4.1** die folgenden beiden Varianten miteinander verglichen (Abbildung 5-1):

- ***spatial***: Die Stimme des Co-Watchers kommt aus der Richtung, in die dieser schaut.
- ***nonspatial***: Die Richtung des Klangs hängt nicht von der Blickrichtung ab.



(a) *spatial*: Die Tonquelle befindet sich in der Mitte des FoV des Sprechers (fromFocus)

(b) *nonspatial*: Der Ton hängt nicht von der Blickrichtung ab.

Abbildung 5-1: Zwei Voice-Chat Methoden: Die Personen schauen sich gemeinsam einen Film über HMD an. Die blauen Symbole gehören zur blauen Person, die orangefarbenen Symbole zur orangefarbenen Person.

Erkenntnisse

Auch wenn wir keinen signifikanten Unterschied in den Antworten des Presence-Fragebogens feststellen konnten, gaben mehrere Teilnehmer in den Kommentaren an, dass die Methode des räumlichen Voice-Chats die räumliche Erfahrung verbessert. Die *spatial*-Methode wurde als komfortabler empfunden und erzeugte mehr Verbundenheit.

5.1.3.2 Nutzerstudie: Emotionszustände senden: *smiley*-Methode, *photo*-Methode

Die Mitteilung emotionaler Zustände kann durch Anzeigen von Symbolen für die Emotionen des Co-Watchers umgesetzt werden (Abbildung 5-2). In unserer **Studie 7.4.1** nutzten wir Smileys (*smiley*-Methode, Abbildung 5-2a) und Fotos von Gesichtern mit verschiedenen Ausdrucksformen (*photo*-Methode, Abbildung 5-2b).

In der **Studie 7.4.1** wurde untersucht, ob diese Methoden das Zusammengehörigkeitsgefühl beeinflussen und welche der beiden Methoden von den Teilnehmern bevorzugt wird.



Abbildung 5-2: Verwendung von Smileys oder Fotos zur Information über die emotionalen Zustände des Co-Watchers.

Erkenntnisse

Die Studienteilnehmer bevorzugten die *smiley*-Methode gegenüber der *photo*-Methode und stuften sie als komfortabler ein. Ein Smiley wirkte im Vergleich zum Foto weniger ablenkend und gab schneller genügend Auskunft über die Gefühle des Co-Watchers. Die Fotos der Gesichter lenkten mehr ab, benötigten mehr Zeit zur Interpretation und erzeugten eine Erwartung an die Kommunikation. Die Teilnehmer wünschten sich jedoch eine größere Auswahl an Smileys. Da wir das Versenden von Smileys ohne Auswahlmenü so einfach wie möglich gestalten wollten, hatten wir uns für die Nutzerstudie dazu entschieden, nur vier Smileys zu verwenden, da dieses mit den vorhandenen Controller-Buttons realisiert werden konnte. Wird der Controller durch Gesten als Eingabeinstrument ersetzt, könnte die Anzahl erhöht werden.

5.1.3.3 Nutzerstudie: Viewport-Indikation: *PiP*-Methode, *bar*-Methode

Das FoV des Co-Watchers zu kennen, ist für die Kommunikation beim Social Viewing wichtig. Eine Möglichkeit der Visualisierung ist die *frame*-Methode, die das FoV des Co-Watchers einrahmt (Nguyen u. a., 2017). Dadurch ist dieses FoV erkennbar, wenn sich die FoVs der Betrachter überlappen (Abbildung 5-3a). Befindet sich das FoV des Co-Watchers außerhalb des Displays, kann ein Pfeil zum Auffinden verwendet werden. Diese Methode eignet sich für Arbeitsaufgaben wie das Bearbeiten von Filmen (Nguyen u. a., 2017).

Eine andere Möglichkeit bietet die *PiP*-Technik (Picture-in-Picture), welche auch in **Studie 7.1.3** als Guiding Methode implementiert wurde. Für die Viewport-Indikation zeigt ein kleines Videofenster das FoV des Co-Watchers an (Abbildung 5-3b). Dies hat den Vorteil, dass das FoV des anderen unabhängig von der eigenen Blickrichtung immer sichtbar ist und man den Bildinhalt sehen kann, ohne sich umzudrehen, aber den Nachteil, dass ein größerer Bereich des Filmbildes abgedeckt wird. Bei Verwendung dieser Methode sollte die Möglichkeit des temporären Abschaltens bestehen. Um das Auffinden des FoV des Co-Watchers zu erleichtern, kann das PiP-Fenster auf der Seite des Displays platziert werden, welche näher am Co-Watcher-FoV liegt.

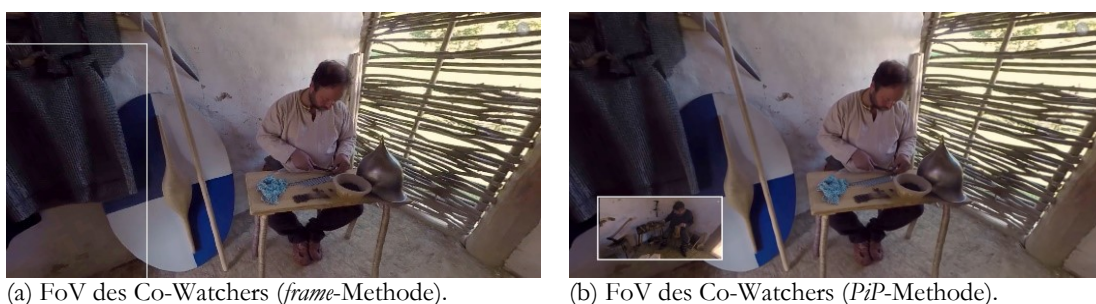


Abbildung 5-3: Ansätze um den Viewport des Co-Watcher zu zeigen. In unserer Studie wurde die *PiP*-Methode verwendet.

Beide bisher genannten Methoden zeigen das genaue FoV des Co-Watchers. Eine weitere Option wäre, nur die Blickrichtung des Co-Watchers und nicht das genaue FoV anzuzeigen. Methoden, die von Segelflugzeugen für Kollisionsvermeidungssysteme genutzt werden, können dafür angewendet werden. Solche Systeme zeigen an, aus welcher Richtung ein anderes Segelflugzeug kommt. Wir implementierten sie in **Studie 7.4.1**, um die Blickrichtung des Co-Watchers darzustellen. Zwei Beispiele sind in Abbildung 5-4a illustriert: Der Balken in Abbildung 5-4a am unteren Rand zeigt an, ob sich das FoV auf der rechten oder der linken Seite befindet (*bar-Methode*). Der Balken rechts visualisiert, ob das FoV höher oder niedriger als die eigene Blickrichtung liegt. Eine Variante ist in Abbildung 5-4b zu sehen, in welcher die Richtung durch einen Kreis und die Höhe durch einen Balken angezeigt wird (*radar-Methode*).



(a) Der Balken unten zeigt an, ob die Blickrichtung des Co-Watchers rechts / links liegt, der Balken rechts, ob diese höher / niedriger ist (*bar-Methode*)



(b) Der Balken rechts zeigt an, ob die Blickrichtung des Co-Watchers höher oder niedriger ist, der Kreis die Richtung (*radar-Methode*).

Abbildung 5-4: Kollisionsvermeidungsmethoden von Segelflugzeugen werden genutzt, um das FoV des Co-Watchers anzuzeigen. In der Studie wurde die *bar-Methode* verwendet.

Nach einer informellen Vorstudie zur Auswahl der Methoden wurden die *PiP*- und die *bar-Methode* in einer Hauptstudie miteinander verglichen

Erkenntnisse

Unsere Studie zeigte keine signifikanten Unterschiede zwischen den Methoden bezüglich der Benutzerfreundlichkeit. In Bezug auf die Zusammengehörigkeit wurde die *PiP*-Methode jedoch von mehr Teilnehmern bevorzugt.

5.1.3.4 Nutzerstudie: Video-Chat: *front-Methode*, *side-Methode*

In den vorherigen Ansätzen können sich die Zuschauer nicht sehen. Um dies zu ermöglichen, wurde ein *PiP*-Video-Chat-Fenster implementiert. Abbildung 5-5 zeigt die zwei Positionen des Fensters, die in **Studie 7.4.1** miteinander verglichen wurden. Bei der Benutzung der *front-Methode* wird die Vorderansicht des Co-Watchers in der unteren Mitte des Bildschirms angezeigt (Abbildung 5-5a), auch wenn der Betrachter den Kopf dreht (screen-referenziert). Die *side-Methode* ist der Situation sehr ähnlich, in der ein Film gemeinsam im Kino oder Fernsehen gesehen wird: Das Chat-Fenster befindet sich world-referenziert an einer fixen Position seitlich des Betrachters (Abbildung 5-5b). Der Zuschauer kann den Co-Watcher nur sehen, wenn er in diese Richtung schaut.

Da beide Betrachter HMDs nutzen, ist ein großer Teil des Gesichts dadurch verdeckt. Diese Tatsache kann die natürliche Kommunikation beeinflussen. Methoden wie „Headset Removal“ (Burgos-Artizzu u. a., 2015; Google Research and Daydream Labs, 2017; Thies u. a., 2016) können dieses Problem möglicherweise in Zukunft lösen. Dies wurde allerdings in unseren Studien nicht näher untersucht.



front-Methode: Das Chat-Fenster ist vor dem Betrachter positioniert (screen-referenziert). Der Co-Watcher ist immer im FoV zentriert.



side-Methode: Das Chat-Fenster ist an der Seite positioniert (world-referenziert). Der Co-Watcher ist zu sehen, wenn in diese Richtung geschaut wird.

Abbildung 5-5: Zwei Methoden zum Anzeigen des Co-Watchers.

In **Studie 7.4.1** wurden die beiden unterschiedlichen Positionen des Chat-Fensters untersucht und die Methoden miteinander verglichen (Abbildung 5-5).

Erkenntnisse

In unseren Tests wurde keine der Video-Chat Methoden von den Teilnehmern bevorzugt und es gab keine signifikanten Unterschiede bezüglich der Benutzbarkeit. Allerdings wies die *side*-Methode für das Item „*Ich weiß, was der andere empfindet*“ einen signifikant höheren Score auf als die *front*-Methode. In der *front*-Methode ist der Mitbeobachter zwar immer sichtbar, das HMD bedeckte jedoch einen großen Teil des Gesichts. Das HMD ist in der *side*-Methode weniger präsent, der Co-Watcher allerdings nur sichtbar, wenn der Betrachter in dessen Richtung schaut. Der Video-Chat wurde in unserer Studie ohne Sprache getestet, da wir nur an der Auswirkung des visuellen Inhalts interessiert waren, um die beiden Video-Chat Methoden zu vergleichen. In einer realen Anwendung sollte diese Komponente durch eine Audiokomponente ergänzt werden.

Der Hintergrund der Personen wurde in den Studien nicht entfernt, damit sich die Chat-Bilder gut vom Filmbild absetzen. Für andere Anwendungen, wie Videokonferenzen, kann der Hintergrund durch den Raum des Co-Watchers ersetzt werden, um das Gefühl zu erzeugen, sich im gleichen Raum zu befinden (Gunkel u. a., 2018). Aber auch für CVR sollte dieser Ansatz untersucht werden.

5.1.3.5 Vergleich der Komponenten

In der **Studie 7.4.1** wurden folgende Methoden getestet:

Komponente	Methode 1	Methode 2
Voice-Chat	spatial	nonspatial
Emotions	smiley	photo
Viewport-Indikation	PiP	bar
Video-Chat	front	side

Es ist nicht nur wichtig für jede Komponente die beste Methode zu finden, sondern auch die Komponenten untereinander in Relation zu setzen. Um zu untersuchen, welche Komponenten für das Zusammengehörigkeitsgefühl von Bedeutung sind, wurden die bevorzugten Methoden jeder Komponente miteinander verglichen:

- Voice-Chat: *spatial*-Methode
- Video-Chat: *front*-Methode
- Viewport: *PiP*-Methode
- Senden von Emotionszuständen: *smiley*-Methode

Die in 7.4.1.4 detailliert beschriebenen Ergebnisse werden im Folgenden kurz zusammengefasst:

In unseren Studien war die *smiley*-Methode neben dem *spatial Voice-Chat* für das Zusammengehörigkeitsgefühl am wichtigsten, gefolgt vom Video-Chat. Für die Teilnehmer war die Kenntnis des FoVs des Co-Watchers von geringerer Bedeutung.

Das *Senden von Emotionszuständen* erreichte hohe Scores für das Zusammengehörigkeitsgefühl. Die Teilnehmer hatten das Gefühl zu wissen, was der andere fühlt und fühlten sich mehr als Teil der Gemeinschaft als bei der Anzeige des Viewports.

Voice-Chat war für die Teilnehmer ebenfalls wichtig, um ein Gefühl der Zusammengehörigkeit zu erzeugen. Nicht nur gesprochene Äußerungen sind relevant, auch ein Lachen oder Seufzen zeigt die Anwesenheit einer anderen Person.

Die *Viewport-Indikation* war die Komponente mit der niedrigsten Bewertung bezüglich der Zusammengehörigkeit. Für die soziale Presence scheint es nicht entscheidend zu sein, die genaue Blickrichtung des Partners zu kennen. Da die gleichzeitige Anwendung aller Komponenten zu einer Überfüllung der Anzeige und zu einer Überlastung der Beobachter führen kann, ist diese Komponente am ehesten verzichtbar, um weggelassen zu werden. Sie kann jedoch wichtig sein, wenn gemeinsam ein sphärischer Film analysiert wird, man daran arbeitet oder Informationen darüber austauscht (Nguyen u. a., 2017).

Der *Video-Chat* wurde in unserer Studie ohne gesprochene Sprache getestet, da wir an den Auswirkungen des visuellen Inhalts interessiert waren, um die beiden Video-Chat-Methoden zu vergleichen. In einer realen Anwendung sollte diese Komponente durch die auditive Komponente ergänzt werden womit möglicherweise ein höheres Zusammengehörigkeitsgefühl erreicht werden kann.

Aus der Multimedia-Psychologie ist bekannt, dass Kombinationen visueller und akustischer Informationen leichter zu verstehen sind als Kombinationen zweier visueller Informationen (Moreno und Mayer, 1999). Daher sollte nur eine Methode pro Sinn verwendet werden. Voice-Chat kann durch Video-Chat oder Smileys erweitert werden. Welche der Komponenten empfehlenswert ist, wird auch von der Art des Films und der Erwartung des Zuschauers beeinflusst (Unterhaltung, Information oder Lernen). Was als nützlich und hilfreich erachtet wird, hängt des Weiteren davon ab, wie viel Dialog mit dem Partner gewünscht wird.

Erkenntnisse

Zusammenfassend wurden die folgenden Schlüsse gezogen:

- *Voice-Chat* und das *Senden von Emotionszuständen* haben die größte Bedeutung für das Gefühl der Zusammengehörigkeit.
- Die genaue *Kenntnis des FoV* des Co-Watchers ist nicht entscheidend und kann durch andere Methoden ersetzt werden (z.B. Ton aus RoI-Richtung).
- Für das *Senden von Emotionszuständen* sollte nach neuen Interaktionstechniken gesucht werden, so dass mehr Typen unterstützt werden können.

5.1.4 Folgerungen für die Herausforderungen

Im Folgenden wird beschrieben, wie sich die untersuchten Methoden auf jede der Herausforderung auswirken. Tabelle 5-1 gibt einen Überblick, welche der einzelnen getesteten Komponenten Lösungsansätze für die in 5.1.2 identifizierten Herausforderungen bieten. Für einige der Herausforderungen sind weitere Konzepte erforderlich, die in Abschnitt 5.1.6 diskutiert werden.

	Voice-Chat	Video-Chat	FoV Anzeige	Emotionszustände
Kommunikation	✓	✓		✓
Kenntnis des FoVs	+		✓	+
Soziale Presence	✓	✓		✓
Barrierefreiheit	+	+		+
Interaktionstechniken	✓	✓	✓	+
Synchronisation				
Multiuser-Umgebungen	+			+

Tabelle 5-1: Die Tabelle zeigt, wie die Ansätze den Herausforderungen entsprechen. ✓: der Ansatz erfüllt die Herausforderung als Ergebnis der Benutzerstudie. +: Der Ansatz sollte angepasst und neu untersucht werden.

Kommunikation

Die Kommunikation beim gemeinsamen Anschauen eines Films kann auf verschiedene Arten erfolgen: verbale Kommunikation per Voice-Chat, visuelle Kommunikation per Video-Chat oder Senden von Bildern (Smileys, Fotos). Jede der visuellen Kommunikationsmethoden kann durch Voice-Chat ergänzt werden. Basierend auf den Ergebnissen unserer Untersuchungen empfehlen wir eine Kombination aus räumlichem Voice-Chat und dem Senden von Smileys. Dabei könnte der Smiley in der Nähe der RoI positioniert werden. Auch wenn sich diese außerhalb des FoV des Co-Watchers befindet, kann die Position gefunden werden, da sie die Quelle des räumlichen Voice-Chats ist.

Kenntnis des Viewports

Die Viewport-Indikation weist im Vergleich zu den anderen Ansätzen die niedrigsten Werte für die soziale Presence auf. Das FoV des anderen zu kennen, scheint für das Miteinander nicht entscheidend zu sein. Dies kann jedoch auch durch die von uns implementierten Methoden verursacht worden sein. Beide Methoden benötigen Zeit zum Eingewöhnen. Demgegenüber wurde der räumliche Voice-Chat, der auch Information zur Blickrichtung des anderen gibt, von den meisten Teilnehmern bevorzugt. Ähnliches wäre mit der *smiley*-Methode möglich, wenn die Smileys in Blickrichtung des Senders positioniert wären. Dies würde jedoch eine zusätzliche Lenkungsmethode erfordern, falls sich dieser Bereich nicht im FoV des Empfängers befindet. Eine weitere Möglichkeit wäre, die *smiley*-Methode durch räumlichen Voice-Chat zu ergänzen. Für die Viewport-Awareness empfehlen wir die gleiche Methodenkombination wie für die Kommunikation: räumlichen Voice-Chat mit dem Senden von Smileys in der aktuellen Blickrichtung.

Soziale Presence

Alle Methoden unserer Nutzerstudie wurden ausgewählt, um die soziale Presence zu verbessern. Die Ergebnisse der Nutzerstudie zeigen, dass das Senden von Emotionsinformationen den größten Einfluss auf das Zusammengehörigkeitsgefühl hat. Die Kenntnis des FoV ergab den niedrigsten Score. Für die soziale Presence ist es offenbar wichtiger zu wissen, was die andere Person fühlt, als zu wissen, was sie gerade sieht. Voice-Chat ist ebenfalls wichtig für die soziale Presence – wichtiger als Video-Chat. Dies kann durch die Tatsache beeinflusst worden sein, dass das HMD während des Video-Chats einen großen Teil des Gesichts bedeckt. Darüber hinaus wirkt der Video-Chat während eines Filmerlebnisses ablenkend. Auch für die soziale Presence empfehlen wir einen räumlichen Voice-Chat in Kombination mit Smileys.

Barrierefreiheit

Personen, die eine Social Viewing Applikation für CVR verwenden, haben unterschiedliche Anforderungen an diese Anwendung. Eine Person mit Sehbehinderung wird den Voice-Chat bevorzugen. Für Menschen mit Hörproblemen kann die *smiley*-Methode oder der Video-Chat wichtiger sein. Außerdem sind einige der implementierten Ansätze geeignet, um Anforderungen der Barrierefreiheit unabhängig von Social Viewing zu erfüllen. Ein CVR-Movie-Player mit den entsprechenden Funktionen kann das Seherlebnis folgendermaßen unterstützen: Voice-Chat kann zur Audiobeschreibung oder zur Erläuterung von Filmen in einfacher Sprache genutzt werden. Ein Gebärdendolmetscher im Video-Chat-Fenster der *front*-Methode kann gehörlose Menschen mit Gebärdensprache unterstützen. Das Senden von Zeichen kann Menschen mit geistiger Behinderung helfen, die Geschichte zu verfolgen. Menschen mit körperlicher Beeinträchtigung (z. B. Rollstuhlfahrer), die sich nur schwer umdrehen können, benötigen die Möglichkeit, die Blickrichtung auf eine andere Weise zu wählen, z. B. mit einem Controller. Solche Methoden werden in den **Studien 7.4.2** und **7.4.3** untersucht und in Abschnitt 5.2.3 erläutert.

Interaktion

Unsere Ansätze sind kopf- und controllerbasierte Methoden. Alle kopfbasierten Eingaben (Blickrichtung für den räumlichen Voice-Chat und FoV-Anzeige) sind natürliche Wege der Interaktion. Für das Senden von Smileys / Fotos wurden Controller verwendet, um sicherzustellen, dass die Aktionen bewusst ausgelöst wurden. Beim Tragen eines HMDs können nur einfache Eingabegeräte verwendet werden. Diese Tatsache begrenzt die Anzahl der verschiedenen Smileys. Die Teilnehmer unserer Nutzerstudie wünschten sich eine größere Auswahl. Dafür würden Tools oder Geräte benötigt, die es ermöglichen, mehrere verschiedene Smileys auszulösen. Gesten könnten ein alternativer Ansatz sein.

Synchronisation

Die Synchronisation hat gut funktioniert. Die Filme wurden jedoch auf dem lokalen Computer bereitgestellt. Nur die Blickrichtung und andere Metadaten wurden übertragen. Dies setzt Vorbereitungen durch Bereitstellung der Filme voraus. Für spontane soziale CVR-Experiences in remote Umgebungen sollten FoV-basierte Bereitstellungs- und Synchronisationstechniken verfügbar sein. In der vorliegenden Arbeit haben wir uns auf Fragen der Mensch-Computer-Interaktion konzentriert und befassen uns nicht eingehender mit diesen technischen Forschungsthemen.

Multiuser-Umgebungen

Alle Ansätze wurden mit zwei Personen getestet. Um sie an mehr Teilnehmer anzupassen, sind Erweiterungen erforderlich, wobei eine Überladung des Displays vermieden werden sollte. Während des Voice-Chats können die Mitglieder anhand ihrer Stimmen unterschieden werden. Manchmal sind die Stimmen jedoch ähnlich und es kann hilfreich sein, visuell zu erfahren, wer spricht, z. B. mit farbigen Lautsprechern oder durch eine Kombination mit der Emotions- oder Video-Chat-Komponente. Mehrere Chat-Fenster würden jedoch den Bildschirm überladen. Gleiches gilt für die Anzeige des FoV. Die *frame*-Methode kann auch für mehr als zwei, jedoch nicht für zu viele Teilnehmer angewendet werden, z.B. durch Nutzen unterschiedlicher Frame-Farben. Smileys könnten ebenfalls leicht durch Farben verschiedenen Betrachtern zugewiesen werden. Nicht immer ist es notwendig und sinnvoll, Informationen aller Gruppenmitglieder gleichzeitig anzuzeigen. Ein zusätzliches Rollenkonzept (siehe Abschnitt 5.1.6), welches die Kanäle und Rechte für jeden Benutzer beschreibt, könnte helfen, die Informationsmenge auf Wesentliches zu reduzieren.

5.1.5 Design-Space für Social Viewing Anwendungen

Die Untersuchung der getesteten Ansätze ergab einen Design-Space für Social Viewing CVR -Applikationen (Tabelle 5-2). Es wurden drei Hauptdimensionen identifiziert: Zuschauer, Nachricht und Gerät. Jede dieser Hauptdimensionen verfügt über Subdimensionen, deren Attribute verschiedene Werte annehmen können. In Tabelle 5-2 sind für jede Dimension die Subdimensionen und mögliche Werte für die Subdimensionen angegeben, wobei die bisher von uns untersuchten Attribute kursiv dargestellt sind. Basierend auf Literaturstudien und unseren Studienergebnissen, sind die genannten Dimensionen die

wichtigsten für Social Viewing CVR-Applikationen. Spezielle Anwendungsfälle erfordern jedoch weitere Überlegungen und Ergänzungen (5.4.1).

	Dimension	Attribute (Optionen)			
Zuschauer	Distanz	<i>co-located</i>	remote		
	Anzahl	<i>Paar</i>	Gruppe (>2)		
	Rolle	<i>gleichrangig</i>	Guide	Follower	Slave
	Beziehung	<i>vertraut</i>	<i>anonym</i>		
Nachricht	Kanal	<i>audio</i>	<i>visuell</i>	haptisch	
	Trigger	<i>kontinuierlich</i>	<i>ausgelöst durch Sender</i>	deaktivierbar durch Empfänger	ausgelöst durch Sensoren
	Type	Information	Emotion	Richtung, Positionen	
	Referenz	screen-referenziert	<i>world-referenziert</i>	switching	
Gerät	Input	<i>Kopf/Blick</i>	Sprache	<i>Controller</i>	Gesten
	Display	<i>HMD</i>	Monitor	mobile Device	CAVE
	Symmetrie	<i>symmetrisch</i>	asymmetrisch		

Tabelle 5-2: Design-Space für Social Viewing in CVR. Für die drei Hauptdimensionen – Zuschauer, Nachricht, Gerät – werden Subdimensionen und mögliche Werte angezeigt. Die implementierten Attribute in den vorgestellten Nutzerstudien sind *kursiv* dargestellt.

Zur Erläuterung der Dimensionen des Design-Spaces werden in diesem Abschnitt die einzelnen Subdimensionen und ihre Attribute beschrieben und die Methoden klassifiziert, die in den Nutzerstudien untersucht wurden. Zusätzlich werden die Studienergebnisse bezüglich der Dimensionen des Design-Spaces diskutiert. Im Anschluss an diesen Abschnitt werden weitere Möglichkeiten und Kombinationen vorgestellt, die sich aus dem Design-Space ergeben.

Zuschauer

Für das Design einer Social CVR-Player-Applikation sind verschiedene Merkmale bezüglich des Zuschauers relevant. Für die Distanz gibt zwei mögliche Szenarien: Die Zuschauer befinden sich im gleichen Raum (*co-located*) oder nehmen an verschiedenen Orten *remote* an der CVR-Applikation teil. Darüber hinaus ist es wichtig, die *Anzahl* der Teilnehmer und ihre *Rollen* zu kennen. Gibt es nur zwei Betrachter (*gepaart*), können beide die gleichen Rechte für die Interaktion mit der Anwendung haben (*gleichrangig*). Dies kann bei *Gruppen* von Teilnehmern allerdings zu Konflikten und zu einer Überladung der Anzeige führen. Das Definieren von Rollen und das Zuweisen von Rechten kann solche Probleme vermeiden. Beispiele für Rollen sind: *Guide* (kann Informationen an alle Mitglieder senden), *Follower* (kann Informationen vom Guide empfangen) und *Slave* (wird zeitlich und in Blickrichtung zum Guide synchronisiert). Weiterhin ist zu berücksichtigen wie gut sich die Teilnehmer kennen (*Beziehung: vertraut* oder *anonym*).

Studien: Unsere Nutzerstudien waren mit Ausnahme der Voice-Chat-Komponente *co-located*. Die Teilnehmer saßen im gleichen Raum, um sie während der Studie gemeinsam beobachten zu können. Die Kommunikation erfolgte allerdings wie in einer *remote* Umgebung über das Netzwerk. Für die Audiokomponente wurde die *remote* Bedingung gewählt, da keine vollständige Geräuschunterdrückung über die Kopfhörer möglich war. In den Studien gab es kein Rollenkonzept und die Teilnehmer waren *gleichrangig*. In allen Tests sahen sich zwei Teilnehmer den Film gemeinsam an (*Anzahl: gepaart*). Einige von ihnen kannten sich, aber nicht alle (*Beziehung: vertraut, anonym*). Wir haben uns bei der Untersuchung nicht auf diesen Aspekt konzentriert und eine Beeinflussung auf die Ergebnisse vermieden,

indem die Teilnehmer hinsichtlich dieser Subdimension zufällig verteilt wurden. Für die Emotionskomponente wurde von den Teilnehmern angemerkt, dass bei anonymen Partnern die *photo*-Methode die Privatsphäre verletzen könnte.

Nachricht

Die zweite Dimension beschreibt die ausgetauschten Benachrichtigungen. Es ist wichtig zu wissen, welche *Kommunikationskanäle* verwendet werden (*akustische, visuelle, haptische*) und wie die Nachricht ausgelöst wird (*Trigger*). Der Design-Space berücksichtigt auch, welcher Art die Benachrichtigungen ist: Informationen über den Filminhalt, eigene Gefühlszustände, Richtungen oder Positionen von Rols. Darüber hinaus ist es relevant, ob die Benachrichtigungen screen- oder world-referenziert sind. Screen-referenzierte Methoden haben den Vorteil, dass sie unabhängig von der Blickrichtung funktionieren. World-referenzierte Methoden sind jedoch oft besser für das VR-Erlebnis, da sie in die VR-Welt integriert sind. Allerdings haben sie den Nachteil, im Fall von visuellen Cues, nicht immer sichtbar zu sein. Dieser Nachteil besteht bei akustischen Cues nicht, da sie unabhängig von der Blickrichtung zu hören sind.

Studien: Wir untersuchten zwei Kanäle: den akustischen (Voice-Chat) und den visuellen (Video-Chat, Emotionsanzeige, Anzeige des FoVs). Dabei wurde festgestellt, dass die akustische Komponente einen großen Einfluss auf die soziale Presence hat. Sie kann eigenständig verwendet oder durch einen visuellen Kanal (z. B. Video-Chat, oder Smileys) ergänzt werden. Die Auswirkungen des *Triggertyps* wurde in unseren Nutzerstudien nicht näher untersucht. Voice-Chat und Video-Chat waren während des gesamten Videoerlebnisses *kontinuierlich* eingeschaltet. Für den Video-Chat wünschten sich einige Teilnehmer jedoch die Möglichkeit, das Chat-Fenster vorübergehend auszuschalten, insbesondere in der Front-Methode. Die gleiche Forderung wurde von den Teilnehmern bzgl. der Anzeige des FoV gestellt. Für den Voice-Chat gab es dazu keine Anmerkungen der Teilnehmer. Für visuelle Methoden wird daher ein Auslösekonzept für Benachrichtigungen benötigt, welches für die akustische Komponente nicht erforderlich ist. Beide Emotionsmethoden (Smileys, Gesichter) wurden vom Absender ausgelöst. Es gab keine Forderungen, es vorübergehend ausschalten zu können. Wir vermuten, dass visuelle kontinuierliche Methoden zu ablenkend sind und Co-Watcher sie bei Bedarf abschalten möchten. In den Studien wurden mehrere *Benachrichtigungstypen* untersucht: Informationen, Emotionen und Blickrichtungen. Die FoV-Anzeige erreichte den niedrigsten Score hinsichtlich der sozialen Presence. Die Teilnehmer zogen es vor, die Blickrichtung des Co-Watchers aus der Richtung der Stimme abzuleiten. Dies ist ein Unterschied zu der in CollaVR untersuchten Kollaboration im Filmschnitt (Nguyen u. a., 2017), bei der der Benutzer das genaue FoV der anderen Person kennen muss. Social Viewing in CVR konzentriert sich auf das Genießen der Anwendung und zu viele zusätzliche Informationen auf dem Bildschirm können dies zerstören. Es ist nicht immer wichtig, die Richtung des anderen Betrachters genau zu kennen. Für den Fall, dass die andere Person etwas über ein Filmdetail per Voice-Chat erzählt, ist es von Vorteil, die Blickrichtung der anderen Person zu kennen, was aber auch aus der Richtung der Stimme herleitbar ist. Ein Ergebnis unserer Studie ist, dass für ein gemeinsames CVR-Erlebnis keine separate Methode für die FoV-Anzeige erforderlich ist. Stattdessen kann die Benachrichtigung über die Betrachtungsrichtung in Kombination mit einer der anderen Methoden gesendet werden, z. B. über räumlichen Voice-Chat oder die Position der Smileys.

Für Methoden in virtuellen Umgebungen ist die *Referenz* ein wichtiges Attribut: In unserer Studie war der spatial-Voice-Chat aus Richtung der Rol world-referenziert. Der Front-Video-Chat war screen-referenziert genau wie die beiden untersuchten Viewport Anzeigen (bar, PiP). Im Gegensatz dazu ist die Framing Methode für die Anzeige des FoVs, die in CollaVR verwendet wird (Nguyen u. a., 2017), world-referenziert. Beide Methoden zum Senden des Emotionsstatus, die in unserer Studie untersucht wurden, waren screen-referenziert. Auf diese Weise waren die Smileys / Bilder immer sichtbar. Werden die Bilder jedoch in der Nähe des FoV des Co-Watchers platziert, ist es einfacher zu erkennen, warum die andere Person Emotionen zeigt. Da das Bild nicht sichtbar wäre, wenn sich das FoV des Senders außerhalb des FoVs des Empfängers befindet, wäre eine Kombination der Techniken eine Alternative: screen-referenziert für die *off-screen* Rol und world-referenziert für eine *on-screen* Rol. Dreht der Betrachter den Kopf, wechselt das Verfahren entsprechend.

Geräte

Als Eingabe für die Kommunikation sind u. a. Controller, Sprache, Gesten und haptische Signale möglich. Die Teilnehmer eine Social CVR-Applikation können HMDs, Monitore, Smartphones oder auch CAVE-Screens als Display nutzen. Die verwendeten Geräte definieren die Symmetrie der Anwendung, die symmetrisch (die Teilnehmer verwenden die gleichen Geräte) oder asymmetrisch (die Teilnehmer verwenden unterschiedliche Geräte) sein kann.

Studien: In unseren Studien wurde das HMD sowohl als Display wie als auch als Eingabegerät für kopfbasierte Methoden verwendet. In der auditiven Komponente hat das HMD die Richtung für den räumlichen Voice-Chat vorgegeben. Zusätzlich wurde ein Controller zum Senden der Emotionsbilder verwendet. Die Anzahl der Smileys war dadurch jedoch auf vier Emotionstypen begrenzt. Bei unserem ersten Ansatz verwendeten alle Teilnehmer das gleiche Display: ein HMD. Alle Tests wurden in symmetrischen Umgebungen durchgeführt.

Unsere Ansätze sind erste Schritte, um Lösungen für soziale Anwendungen in CVR zu finden. Unterstützt durch den beschriebenen Design-Space lassen sich weitere Ansätze finden, die im nächsten Abschnitt diskutiert werden.

5.1.6 Weiterführende Konzepte und Ausblick

Über die Ergebnisse unserer Nutzerstudien hinaus, stellen wir im Folgenden Ideen zu anderen Optionen vor, die in zukünftigen Arbeiten evaluiert werden sollten.

Remote / co-located Umgebungen

Beim Entwerfen einer Social-Viewing-Applikation für CVR muss berücksichtigt werden, ob sich die Zuschauer im gleichen Raum befinden. In diesem Fall können die Co-Watcher direkt per Sprache kommunizieren und Berührungen sind möglich. Die Umgebung (Temperatur, Geruch, Luftstrom, Geräusche im Raum) ist für alle gleich, was das Zusammengehörigkeitsgefühl erhöhen kann. Weitere Untersuchungen sind erforderlich, um festzustellen, ob die Distanz die soziale Presence beeinflusst und ob zusätzliche Komponenten für remote Benutzerszenarien erforderlich sind.

In unserer Nutzerstudie zu Voice-Chat wurde eine remote Umgebung genutzt, um zu vermeiden, dass die Teilnehmer den Co-Watcher aus der realen Welt hören können. Für die Verwendung der räumlichen Voice-Chat Methode in co-located Umgebungen werden Kopfhörer benötigt, mit denen Umgebungsgeräusche vollständig unterdrückt werden können. Allerdings ist es auch möglich, dass in co-located Umgebungen die Stimme aus der realen Welt für die soziale Presence bevorzugt wird. Dies sollte in zukünftigen Arbeiten geprüft werden.

In unseren Ansätzen haben wir nicht untersucht, ob Zuschauer, die sich bereits kennen, andere Optionen oder Parameter benötigen als solche, die sich fremd sind. Beispiele für zu untersuchende Parameter sind die Position und Entfernung des Video-Chat-Fensters. Die Kommunikation für anonyme Betrachter wird dadurch beeinflusst, ob die Anonymität gewahrt bleiben soll. Solche Tatsachen entscheiden die Art der Verwendung von Fotos oder Videos.

Asymmetrische Umgebungen

Co-Watcher in Social-Viewing-Szenarien verfügen nicht unbedingt über die gleiche Hardware. Beispielsweise können sich FoVs unterscheiden, da unterschiedliche HMDs unterschiedliche FoV-Größen aufweisen. Zusätzlich sollte Social Viewing auch für Teilnehmer möglich sein, die Desktops oder begehbare Systeme verwenden. Gugenheimer et al. (Gugenheimer u. a., 2017) implementierten ShareVR, mit dem Benutzer in der realen Welt und Benutzer in einer virtuellen Welt miteinander interagieren können. Sie untersuchten die Asymmetrie in Visualisierung und Interaktion. Die abgeleiteten Richtlinien sollten für CVR verifiziert werden.

Rollenkonzept

Schauen sich mehrere Teilnehmer gemeinsam eine CVR-Applikation an, kann das Anzeigen aller aktuellen FoVs problematisch sein, da dies zu überlasteten Displays führt und sich negativ auf das Seherlebnis auswirkt. Darüber hinaus sollte die Navigationssteuerung nicht für alle Benutzer gleichzeitig zugelassen werden, um chaotische Situationen zu verhindern. Rollenzuweisungen (z. B. Guide, Follower) sowie Methoden zur dynamischen Erteilung von Berechtigungen sind in solchen Fällen erforderlich.

Zwei Hauptszenarien sind denkbar: *non-guided* und *guided*: Der *non-guided* Ansatz beruht darauf, dass allen Zuschauern die gleichen Rollen und Berechtigungen zugewiesen werden. Dieses Szenario eignet sich für zwei Benutzer einer VR-Applikation. Es kann jedoch zu Konflikten bei mehr Teilnehmern und aktiver Kommunikation führen. In solchen Fällen sollte ein *guided* Ansatz in Betracht gezogen werden, in dem den Nutzern Rollen zugeteilt sind: Der *Guide* (z. B. der Kursleiter in Lernszenarien) dient als Referenz für die Synchronisation und ist der einzige Teilnehmer mit aktivierten Steuerungsfunktionen. Ein *Slave*-Modus, bei dem der Follower in Zeit und Blickrichtung mit dem Guide synchronisiert ist, führt allerdings zu Simulator-Sickness (Nguyen u. a., 2017). Ein Slave-Modus kann jedoch in asymmetrischen Umgebungen mit Nicht-VR-Teilnehmern (z. B. Desktop-Nutzer) angewendet werden. Um interaktivere und flexiblere Sitzungen zu ermöglichen, können die Rollen von Guides und Followern dynamisch änderbar sein.

Sensor-Benachrichtigungen

Neben den von den Kommunikationspartnern direkt gesendeten Mitteilungen sind auch andere Arten von Nachrichten vorstellbar. Das Informieren des Partners über die Herzfrequenz oder andere Sensorinformationen könnte ein neuer Ansatz zum Teilen von Emotionen sein (Hassib u. a., 2017). Solche Informationen können jederzeit sichtbar sein (kontinuierlich), vom System ausgelöst werden oder nur Werte über / unter einem bestimmten Schwellenwert anzeigen.

Gesten als Eingabe

Sowohl für die Kommunikation als auch für die Navigation werden einfache Eingabemethoden benötigt. Wir glauben, dass grafische Bedienelemente auf dem Display oder die Bedienung über Sprache das Seherlebnis stören können. Durch Kopf- und Blickbewegungen sind Regionen in der virtuellen Welt auf natürliche Weise adressierbar. Sie sind jedoch nicht für alle Interaktionen ausreichend. Um beispielsweise aus verschiedenen Smileys auszuwählen, ist ein einfacher Auswahlmechanismus erforderlich, der außer über Controller auch durch Gesten realisiert werden kann. Die Verwendung leicht zu merkender Gesten (Abbildung 5-6) würde die Anzahl der Eingabemöglichkeiten erhöhen und weitere Bildtypen ermöglichen. Gesten sind eine natürliche Interaktionsmethode in VR (O'Hagan u. a., 2002), deren Erkennung aber noch nicht immer die notwendige Genauigkeit erreicht (Pakkanen u. a., 2017). Die Genauigkeit von Gestenerkennung wird sich in nächster Zeit weiter verbessern, so dass gestenbasierte Methoden einer erneuten Überprüfung unterzogen werden sollten. Da das Eingabegerät nicht im Fokus unserer Untersuchungen stand, haben wir einen Controller eingesetzt, um technische Probleme zu vermeiden. Wir vermuten jedoch, dass Gesten den Controller ersetzen können.



Abbildung 5-6: Senden von Smileys mit Gesten. Zwei Beispiele für einfache Smileys und die entsprechende Geste.

5.2 Komfortable CVR-Anwendungen

Für die Demonstration von CVR-Applikationen in öffentlichen Räumen wie Filmfestivals oder Messen werden üblicherweise Drehstühle verwendet. Zu Hause oder unterwegs sitzt der Zuschauer dagegen meist auf einem festen Stuhl oder einem Sofa (Hong u. a., 2018). Auf diese Weise ist es schwieriger, den

gesamten 360°-Bereich ohne zusätzliche Techniken zu erkunden. In diesem Abschnitt werden mehrere Ansätze vorgestellt, die es erleichtern, CVR-Filme unter solchen Bedingungen zu schauen. Des Weiteren wird untersucht, ob der gesamte 360°-Bereich für ein immersives CVR-Erlebnis erforderlich ist.

In CVR wählt der Betrachter seinen Bildausschnitt durch Ändern der Blickrichtung selbst aus. Es kann jedoch anstrengend sein, sich die ganze Zeit zu drehen, insbesondere bei längeren Filmen. Auch im realen Leben schauen wir uns nicht ständig um. Meist reicht eine Kopfbewegung aus. Darüber hinaus gibt es in CVR die Befürchtung, etwas zu verpassen, da der Film auch in den Bereichen hinter dem Zuschauer fortgesetzt wird. Diese Fear of Missing Out (FoMo) kann negative Gefühle hervorrufen (Przybylski u. a., 2013) aber durch Guiding-Methoden reduziert werden (Rothe u. a., 2019a).

2017 wurde von Google das Videoformat VR180 eingeführt, das auf dem Standard Spherical Video Metadata V2 basiert (Whitwam, 2018) und von YouTube unterstützt wird. Seitdem sind mehrere 180°-Kameras auf den Markt gekommen. Wir haben jedoch keine Analysen darüber gefunden, ob der reduzierte Bereich die Presence verringert oder in anderer Weise das VR-Erlebnis beeinflusst. Dies soll in diesem Abschnitt untersucht werden. Zusätzlich wird getestet, ob andere Bereichsgrößen besser geeignet sind.

5.2.1 Theoretischer Hintergrund

Human Factors

Menschen sind in der Lage ihren Kopf horizontal um etwa 60° in beide Richtungen zu drehen (Dreyfuss, 1967), komfortable Kopfbewegungen sind dagegen für die meisten Personen bis zu 45° möglich. In einem Bereich zwischen 60° und 120° können Menschen Farben erkennen (Dreyfuss, 1967). Bei Verwendung eines HMDs mit kleinerem FoV ist die Sicht einer Person eingeschränkt. In **Studie 7.4.4** wurde eine Samsung Gear (SM-R323) mit einem FoV von 101° verwendet. Unter Berücksichtigung dieser Größe und eines Kopfbewegungswinkels von 120° (60° in beide Richtungen) kann der Betrachter einen Bereich von etwa 221° in einer festen Sitzhaltung inspizieren (Abbildung 5-7).

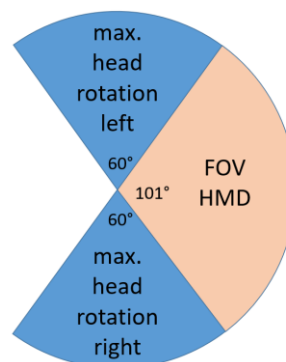


Abbildung 5-7: Bereich, der in einer festen Sitzhaltung gesehen werden kann.

Engagement

Die meisten Menschen schauen sich Filme an, um sie passiv zu genießen und zu entspannen. Im Gegensatz dazu erfordert Gaming mehr Aktivität und Engagement, um Ziele zu erreichen (Dorta u. a., 2016). Durch interaktives Storytelling rücken Filme und Games näher zusammen (Dorta u. a., 2016; Vosmeer und Schouten, 2014). Vosmeer und Schouten (Vosmeer und Schouten, 2014) untersuchten verschiedene Engagement-Stile in Games und Filmen. Betrachter von Filmen erwarten, sich ohne Interaktivität zurücklehnen zu können (*lean-back*), während der Benutzer eines Videospiels mit dem Inhalt interagiert (*lean-forward*). Für den Engagement-Stil in CVR führten Vosmeer et al. den Begriff *lean-in* ein: Der Betrachter schaut sich um und wählt dabei das Sichtfeld selbst aus, so dass es eine Art von Interaktion gibt.

Rotationsmedium

Das Engagement des Betrachters kann durch verschiedene Aspekte beeinflusst werden, unter anderem durch den Typ des Stuhles. In den Experimenten von Hong et al. (Hong u. a., 2018) wurden drei Stuhltypen verglichen: feststehend, halb drehbar und voll drehbar. Bei mehreren abhängigen Variablen (Recall-Rate, allgemeines Engagement und Simulator-Sickness) gab es keine signifikanten Unterschiede zwischen den Testbedingungen. Im Falle eines festen Stuhles nahm jedoch die räumliche Orientierung ab und die FoMO zu, wenn die Hauptfiguren um den Betrachter herum auftraten. Außerdem erkundeten die Teilnehmer die Szenerie intensiver, wenn sie den Drehstuhl benutzten.

Gugenheimer et al. (Gugenheimer u. a., 2016b) entwickelten einen motorisierten Drehstuhl (SwiVRChair), der den Betrachter dabei unterstützt, die Geschichte zu verfolgen, indem er ihn automatisch in Richtung der RoI dreht. Auf diese Weise benötigten die Teilnehmer weniger Kopfbewegungen, um das VR-Erlebnis in einer lean-back-Weise genießen zu können. Die Simulator-Sickness war in diesem Experiment sehr gering, was dadurch verursacht sein kann, dass die Zuschauer real gedreht wurden und so die Rotation in der virtuellen Welt mit der Rotation in der realen Welt übereinstimmte. Eine Ursache für Simulator-Sickness ist die Diskrepanz zwischen Bewegungen in realer und virtueller Welt (Davis u. a., 2014; Kolasinski, 1995). Es gibt keine einheitliche Meinung darüber, ob das Drehen einer Szene zu Simulator-Sickness führt (Y.-C. Lin u. a., 2017).

Forcierte Methoden können angewendet werden, um den Benutzer zur RoI zu führen (Gugenheimer u. a., 2016b; Y.-C. Lin u. a., 2017; Stratmann u. a., 2018). Dafür gibt es verschiedene Möglichkeiten. Einerseits kann der Benutzer gedreht werden, wie in SwiVRChair (Gugenheimer u. a., 2016b), was den Vorteil hat, dass der Betrachter die Drehbewegung spüren kann. Andererseits kann die VR-Welt / der Film gedreht werden (Y.-C. Lin u. a., 2017). Lin et al. (Y.-C. Lin u. a., 2017) verglichen forcierte Rotation (dort Autopilot genannt) mit manueller Rotation, bei der ein Pfeil die Richtung der RoI anzeigt. Die Ergebnisse waren abhängig von der Art des Films, es wurde jedoch keine allgemeine höhere Simulator-Sickness für die forcierte Rotation beobachtet.

Verstärkung der Kopfdrotation (rotational gain / amplified head rotation)

Drehungen der virtuellen Welt, die nicht vollständig synchron zu den Bewegungen des Benutzers sind, werden für *Redirect Walking* Techniken verwendet (Razzaque u. a., 2001; Steinicke u. a., 2008). Bei diesen Techniken wird die VR-Welt in einer für den Benutzer unbemerkten Weise geändert, so dass andere Wege in der virtuellen Welt zurückgelegt werden, als in der realen Welt. Dadurch kann der real getrackte Bereich in der virtuellen Welt einen größeren Bereich abdecken. Steinicke et al. (Steinicke u. a., 2010) untersuchten die Schwellwerte für diese Technik.

Jaekl et al. (Jaekl u. a., 2005) führten eine verstärkte Kopfdrehung für HMDs ein (rotation gain). In ihren Experimenten bestimmten die Teilnehmer den gain-Faktor, für den die virtuelle Welt in Bezug auf ihre eigenen Kopfbewegungen stabil ist. Die identifizierten Skalierungsfaktoren lagen zwischen 0,88 und 1,33 (Gieren, yaw) bzw. zwischen 0,11 und 1,66 (Neigen, pitch). Diese Ergebnisse entsprechen den Ergebnissen von Steinicke et al. (Steinicke u. a., 2010) (Gieren: 0,83-1,34). Skalierungsfaktoren in diesem Bereich werden als kompatibel mit der Wahrnehmung toleriert, sich in einer stabilen Welt zu bewegen, und Benutzer sind sich der Manipulation nicht bewusst. Jay und Hubbard (Jay und Hubbard, 2003) erhöhten den gain-Faktor und stellten bei visuellen Suchaufgaben signifikante Performanceverbesserungen fest. Unter Verwendung eines Skalierungsfaktors von 2 fühlten sich die Teilnehmer wohl und zogen sogar die gain-Methode der Baseline ohne Verstärkung vor.

Verstärkte Kopfbewegungen eignen sich gut für Beobachtungsaufgaben. Es ist jedoch schwierig, damit in einer VR-Welt zu interagieren, in der die anderen Körperbewegungen nicht auf die gleiche Weise verstärkt werden (Jay und Hubbard, 2003). Dieses Problem ist allerdings für CVR nicht relevant.

Die gain-Methode kann angewendet werden, damit Zuschauer CVR-Filme auf einem festen Stuhl mit weniger körperlicher Bewegung genießen können. Hong und Kim (Hong und Kim, 2016) verglichen dynamische und konstante Rotationsverstärkung. Für die konstante Methode wurde ein Skalierungsfaktor von 1,3 verwendet, für die dynamische Methode zwischen 1,3 und 1,6. Die dynamische Verstärkung

wurde so implementiert, dass die Verstärkung für schnelle Kopfbewegungen höher ist. Beide Methoden funktionierten gut und erzeugten weder Simulator-Sickness noch Einschränkungen im Nutzererlebnis.

Ein weiteres dynamisches gain-Verfahren wurde von Sargunam et al. (Sargunam u. a., 2017) entwickelt, bei dem der gain-Faktor im nahen Bereich geringer ist als in entfernten. Bei dieser Methode wird der gesamte virtuelle 360°-Bereich durch Kopfbewegungen von 90° nach rechts und 90° nach links abgedeckt. In ihren Experimenten bevorzugten die Teilnehmer allerdings die Standardrotation ohne jegliche Verstärkung.

Eine Rotationsverstärkung macht es schwieriger, ein Ziel genau zu fokussieren. Langbehn et al. (Langbehn u. a., 2019) untersuchten verschiedene dynamische Rotationsmethoden, die auf das Ziel ausgerichtet waren. Dabei schnitt die dynamische Verstärkung im Vergleich zur statischen Verstärkung, einer scroll-Methode und der Baseline am besten ab. Da sich die Betrachter in unseren Experimenten ohne jede Aufgabenstellung umsehen, sind solche zielgerichteten dynamischen Methoden nicht anwendbar.

Blickverhalten

1996 untersuchten Pausch et al. (Pausch u. a., 1996), wie die Aufmerksamkeit des Betrachters auf eine gewünschte RoI gelenkt werden kann. Die Position und Ausrichtung des Kopfes wurden aufgezeichnet und herkömmliche Histogramme zur Veranschaulichung der Daten verwendet. Die Ergebnisse zeigen, dass die meisten Menschen ihren Kopf nicht sehr oft drehen, und fast immer in einem Bereich von -60° bis + 60° in horizontaler Richtung und von -30° bis + 20° in vertikaler Richtung blieben. Dies ist konform zu den Forschungen von Sitzmann et al. (Sitzmann u. a., 2018), in denen gezeigt wurde, dass Zuschauer dazu neigen, vor allem den Bereich um den Äquator herum anzuschauen. In **Studie 7.4.4** untersuchten wir, ob der 360°-Bereich in horizontaler Richtung verkleinert werden kann, um das CVR-Erlebnis angenehmer zu gestalten und FoMO zu verringern.

5.2.2 Herausforderungen für angenehme CVR-Applikationen

Aus der Analyse der oben beschriebenen Literatur, den Erfahrungen unserer eigenen Arbeit und den bisher beschriebenen Erkenntnissen, halten wir die folgenden Herausforderungen für den Nutzerkomfort für besonders wichtig, um CVR -Applikationen bequem und auch für längere Zeit verwenden zu können.

Herausforderung 1 – Isolationsreduzierung: Das Problem der Isolierung durch HMDs gehört zu den Grundproblemen in CVR und wir haben ihm deshalb besondere Aufmerksamkeit geschenkt und in Abschnitt 5.1 ausführlich beschrieben.

Herausforderung 2 – Geräte: Ein von den Nutzern oft bemängelter Punkt ist die Unbequemlichkeit der momentan verfügbaren Geräte. Für eine längere Anwendung und entspannte Seherlebnisse sind die Geräte noch zu schwer und unhandlich. Durch neue technische Entwicklungen wurden in den letzten Jahren viele Verbesserungen erzielt: HMDs lassen sich auch kabellos benutzen und sind leichter geworden. Diese technische Herausforderung ist nicht Thema dieser Arbeit, soll aber hier der Vollständigkeit halber erwähnt sein.

Herausforderung 3 – Rotation: Das ständige Umherschauen in einer CVR-Umgebung ist auf Dauer anstrengend und nicht in allen Situationen möglich. Um auch auf einer Zugfahrt oder auf dem Sofa eine CVR-Anwendung genießen zu können, braucht es entweder Inhalte, die ein Umdrehen unnötig machen oder Techniken, die es auch in solchen Situationen ermöglichen, jeden Bildausschnitt zu sehen.

Herausforderung 4 – FoMO: Wie schon in Kapitel 2 erwähnt, belastet viele Zuschauer der Gedanken, dass sie nur einen Teil des Films sehen. Die Befürchtung etwas zu verpassen, verringert das Zuschauererlebnis. Einige Methoden, dem gegenzusteuern, wurden bereits in Kapitel 2 aufgezeigt und werden hier ergänzt.

5.2.3 Methoden zur Gestaltung komfortabler CVR-Applikationen

Die **Studien 7.4.2 - 7.4.4** konzentrierten sich auf die unter 5.2.2 genannte Herausforderung 3 und 4. Im Folgenden werden diese drei Ansätze vorgestellt, um Lösungen für diese Herausforderungen zu finden. Für jeden Ansatz implementierten wir verschiedene Methoden, die in den Studien miteinander verglichen wurden.

5.2.3.1 Immobile Techniken: *gain*-Methode, *var*-Methode, *const*-Methode

Da das Anschauen von Filmen in vielen Fällen eine entspannende Tätigkeit ist, sind Techniken erforderlich, um den gesamten Bereich eines CVR-Films zu erkunden, ohne den Körper ständig zu drehen. Solche Techniken lassen sich nicht nur auf dem Sofa, sondern auch am Schreibtisch, in einem Zug oder an anderen Orten mit eingeschränkter Bewegungsfreiheit anwenden.

Für **Studie 7.4.2** wurden drei Methoden entwickelt, die es ermöglichen, den gesamten 360°-Bereich mittels Controller oder Rotationsverstärkung ohne Körperdrehung anzuschauen (Abbildung 5-8).

Dazu analysierten wir als Erstes die verstärkte Kopfdrotation (Jaekl u. a., 2005). Die Ergebnisse von Jaekl zeigen, dass ein *gain*-Faktor zwischen 0.88 und 1.33 für die Drehung des Kopfes von den Teilnehmern nicht als Modulation wahrgenommen wird. Unter Berücksichtigung eines Bereichs von 120° für maximale Kopfbewegungen und eines HMD-FoVs von 100° ist ein Verstärkungsfaktor von 1.33 nicht ausreichend, um den gesamten 360°-Bereich abzudecken.

In einer informellen Pilotstudie wurde ein *gain*-Faktor von 2.0 ermittelt. Ein *gain*-Faktor von 2 ist für den Betrachter angenehm, auch wenn die Verstärkung bemerkt werden kann. Der Faktor 2 hat einen weiteren Vorteil. Selbst wenn sich Leute vollständig drehen, haben sie immer die gleichen Filmabschnitte an der gleichen Position, der Filminhalt wird bei einer 360°-Drehung genau zweimal dargestellt. Faktoren, die keine ganzen Zahlen darstellen, würden dagegen zu Versetzungen führen. Es wurden die folgenden Methoden miteinander verglichen:

- ***gain***: Die horizontale Kopfdrehung wird um den Faktor 2 verstärkt. Dreht der Nutzer seinen Kopf in der realen Welt um 10°, ändert sich der Blickwinkel in der virtuellen Welt um 20°.
- ***var***: Der Zuschauer kann das Bild drehen, indem die Controller-Taste nach rechts oder links betätigt wird. Je mehr die Taste nach rechts oder links gedrückt wird, desto höher ist die Rotationsgeschwindigkeit bis zu einer maximalen Geschwindigkeit von 60°/s. Kopfbewegungen können zusätzlich zum Inspizieren der Szenerie verwendet werden.
- ***const***: Der Zuschauer kann das Bild mit konstanter Geschwindigkeit von 45°/s drehen, indem die Controller-Taste nach rechts oder links betätigt wird. Die Drehung ist nur in horizontaler Richtung möglich. Kopfbewegungen können zusätzlich zum Inspizieren der Szenerie verwendet werden.



Abbildung 5-8: Verschiedene Techniken für immobiles CVR: links: Verstärkung der Kopfdrehung (*gain*), Mitte: Rotation des Films mit Controller und variabler Geschwindigkeit (*var*), rechts: Rotation des Films mit Controller und konstanter Geschwindigkeit (*const*).

Erkenntnisse

Die Methoden unterschieden sich in Bezug auf Presence, Nutzererfahrung und Simulator Sickness nicht signifikant. Die Ergebnisse zeigen, dass es neben dem Drehstuhl noch andere Methoden gibt, um CVR-Filme zu schauen. Solche Methoden erfordern weniger körperliche Anstrengung und können in einer festen Sitzhaltung genutzt werden. Auch wenn die Verwendung eines Controllers oder einer Verstärkung der Kopfrotation (*gain*-Methode) eine Diskrepanz zwischen Bewegungen in der realen und der virtuellen Welt erzeugen, berichteten in unseren Experimenten die Teilnehmer kaum über Simulator-Sickness, was möglicherweise dadurch beeinflusst ist, dass die Rotation selbst initiiert wurde. Zusammenfassend fanden wir folgende Ergebnisse:

- Die *var*- und die *gain*-Methode wurden von den Zuschauern akzeptiert.
- Die *const*-Methode ergab die niedrigsten Scores. Bei Nutzung eines Controllers sollte die *var*-Methode der *const*-Methode vorgezogen werden.
- Die *gain*-Methode ergab höhere Scores für die Experience als die beiden Controller-Methoden, verursachte jedoch ein wenig mehr Simulator-Sickness.

5.2.3.2 Rotationsmethoden: *fixed*, *swivel*, *controller*

In einer weiteren **Studie (7.4.3)** wurden drei Rotationsmethoden untereinander verglichen: die Rotation des Körpers mittels eines Drehstuhls, die alleinige Bewegung des Kopfes und die Rotation des Films mittels Controller (Abbildung 5-9). Bei allen drei Methoden sind die Kopfbewegungen in der realen und virtuellen Welt synchron, ohne Rotationsverstärkung. Da in **Studie 7.4.2** die *var*-Methode gegenüber der *const*-Methoden von den Nutzern bevorzugt wurde, wird im Weiteren nur noch die *var*-Methode als *controller*-Methode genutzt.

- **swivel:** Der Zuschauer sitzt auf einem Drehstuhl und kann sich in alle Richtungen drehen. Außerdem kann der Kopf bewegt werden, um sich umzusehen.
- **fixed:** Der Zuschauer sitzt auf einem festen Stuhl und kann sich ohne zusätzliche Rotationsverstärkung umsehen, indem der Kopf bewegt wird. Die Teilnehmer wurden darüber informiert, dass die Handlung des Films vor ihnen abläuft und keine für die Geschichte wichtigen Details hinter ihnen stattfinden.
- **controller:** Der Zuschauer kann das Bild drehen, indem die Controller-Taste nach rechts oder links betätigt wird. Je mehr die Taste nach rechts oder links gedrückt wird, desto höher ist die Rotationsgeschwindigkeit bis zu einer maximalen Geschwindigkeit von 60°/s. Kopfbewegungen können zusätzlich zum Inspizieren der Szenerie verwendet werden (entspricht der *var*-Methode aus dem vorigen Abschnitt).

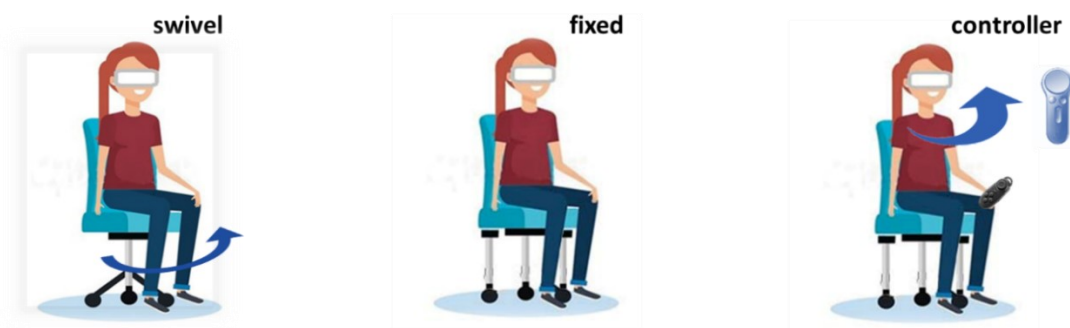


Abbildung 5-9: In **Studie 7.4.3** wurden verschiedene Techniken miteinander verglichen: links: Drehstuhl (*swivel-chair*), Mitte: feste Sitzposition (*fixed*), rechts: Rotation des Films mit Controller (*controller*).

Die 360°-Filme für diese Studie wurden so ausgewählt, dass die Handlung in einem 180°-Bereich ablief.

Erkenntnisse

Die Methoden unterschieden sich in Bezug auf Presence, Erlebnis und Simulator-Sickness nicht signifikant. Diese Ergebnisse legen nahe, dass die Verwendung eines Drehstuhls zum Ansehen von CVR-Filmen durch andere Methoden ersetzt werden kann.

Selbst wenn sich die Handlung der Filme auf einen 225°-Bereich konzentrierte und die Teilnehmer darüber informiert waren, ließ die Kenntnis, dass der Film sie umgab, bei den Zuschauern mit der *fixed*-Methode die Befürchtung aufkommen, etwas zu verpassen. Die Tatsache, dass wir die Rückseite des Films nicht gelöscht hatten, könnte dieses Ergebnis beeinflusst haben. Aus diesem Grund sahen die Teilnehmer sich auch auf dem festen Stuhl um, was unbequem war. Die Teilnehmer bevorzugten in der Studie den Controller und den Drehstuhl, gaben jedoch an, generell Filmen mit weniger als 360° den Vorzug zu geben.

- Neben dem Drehstuhl ist auch ein Controller für die Rotation verwendbar.
- Die Zuschauer wünschen sich Filme, die einen kleineren Bereich als 360° abdecken.

5.2.3.3 Weniger als 360°: 180°, 225°, 360°

Ein Aufnahmebereich von weniger als 360° kann mehrere Vorteile bringen: (1) Die Befürchtung etwas zu verpassen (FoMO) könnte verringert werden. (2) Die Produktion von CVR-Videos wäre einfacher, da bei den Aufnahmen ein freier Bereich für Licht und Filmteam zur Verfügung stehen würde. (3) Die Bildqualität könnte mit der gleichen Datenmenge verbessert werden.

In unserer **Studie 7.4.4** wurde untersucht, wie ein begrenzter Bereich die Presence, die kognitive Belastung und das Seherlebnis beeinflusst. Dafür wurden neben dem 360°-Filmsektor auch 180°- und 225°-Sektoren betrachtet (Abbildung 5-10 und Abbildung 5-11).

Der Bereich von 180° wurde gewählt, da dieser auf YouTube-Plattformen bereits genutzt wird (Rodriguez, 2017). Die Grenzen dieses Bereichs wurden mit einem 4.5° breiten Rand ausgeblendet. Horizontale Kopfbewegungen sind für Menschen in einem Bereich von bis zu 45° in jede Richtung angenehm (Dreyfuss, 1967). Unter Berücksichtigung des 101° FoV der Samsung Gear konnten die Teilnehmer einen Bereich von 191° wahrnehmen und damit die Grenzen des 180°-Sektors erkennen, wenn sie ihren Kopf nach links oder rechts bewegten.

Der 225°-Bereich wurde aufgrund des FoV des Samsung Gear HMDs (101°) und der maximalen Kopfdrehung des Menschen (60° in jede Richtung) gewählt. Die Grenzen wurden mit einem 4.5° breiten Rand ausgeblendet. Bei bequemen Kopfbewegungen (45° in jede Richtung) waren die Grenzen nicht sichtbar. Die Grenzen waren jedoch bei extremen Kopfbewegungen zu sehen.

In beiden Fällen sahen die Teilnehmer den Film in einer festen Sitzposition und bewegten nur den Kopf

Mit der 360°-Methode konnten die Betrachter den gesamten 360°-Bereich mit einem Drehstuhl inspizieren.

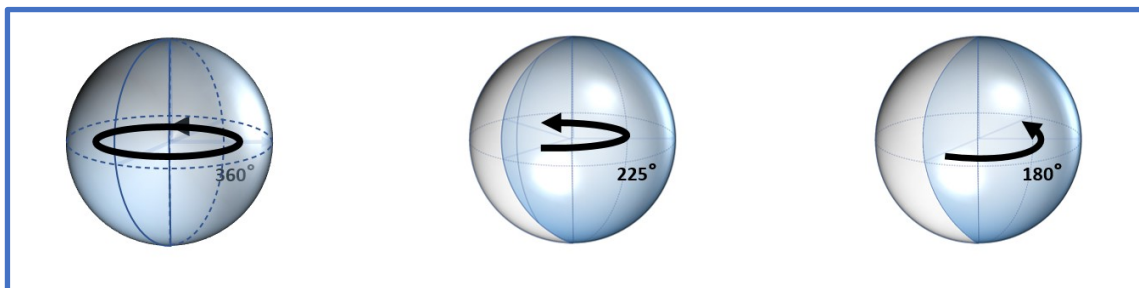


Abbildung 5-10: Die drei Konditionen dargestellt auf der Sphäre.

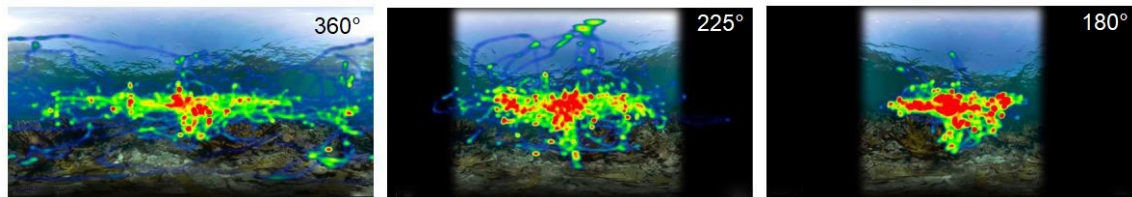


Abbildung 5-11: **Studie 7.4.4** - verschiedene Größen der CVR-Projektionsfläche als äquierektanguläre Projektionen, links: Drehstuhl (360°), Mitte: feste Sitzposition (225°), rechts: feste Sitzposition (180°).

Erkenntnisse

Beim Vergleich von 360° -, 225° - und 180° -Filmen in **Studie 7.4.4** wurde festgestellt, dass der eingeschränkte Betrachtungsbereich die Presence geringfügig verringert. Den Zuschauern war bewusst, dass sich hinter ihnen nichts befindet, auch wenn der Inhalt vor ihnen und die Grenzen in der 225° -Bedingung nur bei extremen Kopfdrehungen sichtbar waren.

- Die Presence wird durch die Verkleinerung der Filmbereiche auf 225° / 180° geringfügig reduziert.
- Zeitdruck, körperliche und geistige Belastung sind im 360° -Fall am höchsten.
- Die Betrachtung des gesamten 360° -Bereichs ist körperlich und geistig anstrengender und erzeugt mehr Zeitdruck und FoMO.
- Die Betrachtung reduzierter Bereiche von 225° oder 180° ist entspannender. Die 225° -Methode wurde der 180° -Methode vorgezogen.

5.2.4 Folgerungen für die Herausforderungen

Zusammenfassend wurden im Abschnitt 5.2 folgende Ergebnisse gefunden:

- In festen Sitzpositionen können mehrere Methoden zum Ansehen von CVR-Videos verwendet werden: Rotation des Films mittels Controller (controller-Methode), Verstärkung der Kopfdrehung (gain-Methode) und Nutzung eines Drehstuhles.
- Durch die Verkleinerung der Filmbereiche auf 225° , bzw. 180° werden Zeitdruck, körperliche und geistige Belastung reduziert und FoMO vermindert. Die Presence nimmt jedoch geringfügig ab.
- Die 225° -Methode wurde der 180° -Methode vorgezogen.

Die vorgestellten Methoden sollen die aktiven Methoden, in denen sich der Betrachter viel umdreht, nicht ersetzen, sondern ergänzen. Abhängig vom Genre des Films, dem Inhalt und dem Wunsch des Zuschauers sollte die geeignetste Methode angewendet werden.

5.3 Barrierefreie CVR- Anwendungen

Fernsehsender und Filmanbieter bemühen sich seit Jahren darum, ihre Medien barrierefrei zu gestalten. Dazu gehören Untertitel für Menschen mit eingeschränktem Hörvermögen, Gebärdensprache für gehörlose Menschen, Audiodeskription für Menschen mit eingeschränktem Sehvermögen und Leichte Sprache (ARD, 2018) für Menschen, für die komplexe Sätze schwer verständlich sind. Für CVR-Anwendungen kommt hinzu, dass auch körperliche Einschränkung den Mediengenuss beeinträchtigen können. Findet man Wege CVR barrierefrei zu gestalten, bieten sich neue Möglichkeiten für ungewohnte Erfahrungen: Plätze, die aus verschiedenen Gründen nicht erreichbar sind, können virtuell besucht werden; man kann einen Eindruck der Unterwasserwelt auch ohne Tauchgang bekommen oder auch ein Museum ohne großen Anreisearaufwand besuchen.

5.3.1 Theoretischer Hintergrund

Barrierefreiheit in VR

Verschiedene Arbeiten setzen sich mit der Barrierefreiheit in VR auseinander (Mott u. a., 2019). Durch VR ist es möglich, Aktivitäten in einer geschützten Umgebung auszuführen, die in der realen Welt schwer oder nicht ausführbar wären (Wilson u. a., 1997). Blinde Menschen können beim Trainieren von Taktiken und bei der Aneignung räumlicher Umgebungskonfigurationen mit akustischen Raumsimulationen in VR unterstützt werden (Picinali u. a., 2014). Zhao et al. entwickelten SeeingVR, ein Set von 14 Tools, aus denen Menschen mit Sehbehinderung verschiedene Werkzeuge entsprechend ihrer Bedürfnisse auswählen, anpassen und kombinieren können. Dadurch ist es möglich die VR-Anwendung für Menschen mit Sehbehinderung durch visuelle und akustische Erweiterungen zu verbessern.

Positionierung von Untertiteln in framed Filmen

Untertitel sind sowohl für Menschen mit Hörbehinderungen wie auch für Nicht-Muttersprachler wichtig. In framed Filmen werden in der Regel *statische* Untertitel verwendet. Diese Untertitel befinden sich meist am unteren Bildrand und ändern ihre Position nicht. Diese Methode wird auch als *center-bottom* bezeichnet (Kurzhaus u. a., 2017). In CVR sind Untertitel, die im unteren Bereich des Displays statisch verknüpft sind, nicht statisch bezüglich des Filmbildes, sondern bewegen sich mit, falls der Betrachter den Kopf dreht. Wie in Kapitel 2 erläutert, werden in VR-Umgebungen solche Objekte als *screen*-referenziert bezeichnet (Shah und Miyake, 2005; Yeh u. a., 1999).

Um Kopf- und Augenbewegungen beim Betrachten traditioneller Filme mit Untertiteln gering zu halten, können *dynamische* Untertitel in der Nähe des Sprechers genutzt werden. Die Position dieser Untertitel ändert sich dynamisch und hängt von der Szene ab. Andere Bezeichnungen für diese Untertitel sind: *speaker following Subtitles* (Kurzhaus u. a., 2017) oder *positioned Subtitles* (Brooks und Armstrong, 2014). In CVR sind sie in der virtuellen Filmwelt platziert und bleiben an ihrem Platz fixiert, wenn der Betrachter den Kopf dreht. Sie sind *world*-referenziert (Shah und Miyake, 2005; Yeh u. a., 1999) und ihre Sichtbarkeit ist abhängig von der Blickrichtung.

In mehreren Studien wurde die Positionierung dynamischer Untertitel in framed Filmen untersucht, mit dem Ziel das Lesen der Untertitel angenehmer zu machen, die Anzahl der Blickpositionswechsel und den Abstand zwischen Rol und Untertiteln zu verringern (Akahori u. a., 2016; Brooks und Armstrong, 2014; Brown u. a., 2015; Hong u. a., 2011; Hu u. a., 2015). Akahori et al. (Akahori u. a., 2016) bestimmten Regionen, in die die Zuschauer blickten, anhand von Eyetracking-Daten eines Films und platzierten die Untertitel im unteren Teil dieses Bereichs, um die Augenbewegungen zu verringern. Die Experimente von Chen et al. (Chen u. a., 2017) zeigten, dass bei dynamischen Untertiteln (Untertitel in Nähe der Sprecher) die Lernenden höhere Punktzahlen erreichen und mehr Interesse am Video zeigen als bei der Verwendung statischer Untertitel. In unserer Arbeit untersuchen wir, ob dynamische / world-referenzierte Untertitel in CVR-Umgebungen anwendbar sind und ob die Methoden Einfluss auf das Seherlebnis haben.

Kurzhaus et al. (Kurzhaus u. a., 2017) verglichen mittige (center-bottom) Untertitel mit dynamischen (speaker-following) Untertiteln in herkömmlichen Videos. Dynamische Untertitel führten zu einer höheren Fixierungshäufigkeit von Rols und reduzierten Sakkadenlängen. Die Teilnehmer hatten den subjektiven Eindruck, den Inhalt mit dynamischen Untertiteln besser zu verstehen. In den Experimenten wurde der Ton stummgeschaltet. Da in CVR Audio ein wichtiges Indiz für Veränderungen in der Szene - auch außerhalb des FoV - ist, haben wir diesen Ansatz nicht übernommen. Stattdessen wurden die gesprochenen Teile durch Reverse-Audio-Filter manipuliert und unverständlich gemacht.

Verschiedene Messmethoden zum Vergleich der kognitiven Belastung mit und ohne Untertitel in herkömmlichen Filmen wurden von Kruger et al. genutzt (Kruger u. a., 2013): Tracking der Pupillendilatation, Elektroenzephalographie (EEG) und Fragebögen. In ihren Experimenten reduzierten gleichsprachige Untertitel im pädagogischen Kontext die kognitive Belastung (gemessen mittels Pupillendurchmesser) und führten zu einem niedrigeren Frustrationsniveau (gemessen mittels EEG).

Brown et al. (Brown u. a., 2015) analysierten die Eyetracking-Daten für Untertitel in framed Videos. Sie stellten fest, dass die Blickmuster von Personen, die dynamische Untertitel betrachten, den Mustern der Baseline ähnlicher waren als bei herkömmlichen Untertiteln. Die meisten Teilnehmer waren konzentrierter und verpassten weniger Inhalte. Einige bevorzugten jedoch traditionelle Untertitel, weil die dynamischen Untertitel das Bild beeinträchtigten. Ein weiterer Nachteil ist die Tatsache, dass dynamische Untertitel für Zuschauer, die keine Untertitel benötigen, störender wirken. Diese Schwäche ist für CVR nicht relevant, da im Gegensatz zu herkömmlichen Videos, bei denen mehrere Personen auf dasselbe Display schauen, jeder Zuschauer für sich entscheiden kann, ob Untertitel gewünscht sind.

Sprecheridentifikation in framed Videos

Ein weiterer Aspekt neben der Positionierung von Untertiteln ist die Identifizierung von Sprechern in Szenen mit mehreren Personen. Das Platzieren der Untertitel in der Nähe des Sprechers ist eine Methode, die zur Lösung des Problems beitragen kann (Vy und Fels, 2010).

Vy und Fels (Vy und Fels, 2010) verglichen Untertitel, denen Sprecheramen vorangestellt sind, mit Untertiteln neben dem Sprecher. In ihren Experimenten fühlten sich die Teilnehmer abgelenkt durch Untertitel, die den Sprechern folgten und dadurch den Ort wechselten. Sprecheramen waren für die meisten Teilnehmer hilfreich, nicht jedoch für gehörlose Zuschauer, die die Stimmen nicht kennen und Personen normalerweise nicht anhand von Namen, sondern anhand von visuellen Merkmalen identifizieren. Andere Methoden sind: unterschiedliche Schriftfarben, Symbole und andere grafische Elemente (King und et al., 1994; Vy und Fels, 2010, 2009).

Das Problem der Sprecheridentifikation ist in Cinematic VR relevanter als in herkömmlichen Videos, da die Personen gleichzeitig im Space vorhanden sind, auch wenn der Betrachter nur einen Teil davon im FoV sieht.

Untertitel in Augmented Reality

CVR hat viele Ähnlichkeiten mit AR, wo die reale Welt den Platz des 360°-Films einnimmt. Peng et al. (Peng u. a., 2018) untersuchten Sprechblasen, die dynamischen Untertiteln sehr ähnlich sind. Ein System mit Echtzeit-Spracherkennung für AR-HMDs wurde entwickelt, um gehörlose und schwerhörige Personen bei Gesprächen zu unterstützen. Die Teilnehmer bevorzugten Sprechblasen gegenüber herkömmlichen Untertiteln. Für unsere Studien haben wir uns entschieden, keine Sprechblasen zu verwenden, da diese einen zu großen Bereich des Films überdecken. Die Position von world-referenzierten Untertiteln in der Nähe des Sprechers und den Sprechblasen dieser AR-Studie ist jedoch gleich. Probleme bei der Sprecheridentifikation und der Einfluss von Untertiteln auf die Presence unterscheiden sich aber in AR und CVR.

Statische Untertitel in CVR

In ihrer Arbeit schlugen Brown et al. (A. Brown u. a., 2017) vier statische Untertitelungsmethoden für CVR vor:

- *120-Grad*: Die Untertitel werden an drei verschiedenen Stellen um den Betrachter (0°, 120°, 240°) und 15° unterhalb der Horizontlinie (0°) platziert.
- *Static-Follow*: Die Untertitel sind vor dem Betrachter fixiert und statisch mit den Kopfbewegungen verbunden. Sie liegen 15° unter der Horizontlinie (0°),
- *Lag-Follow*: Ähnlich der Static-Follow-Methode, jedoch bleiben die Untertitel für kleine Kopfbewegungen (unter 30°) an ihrem Platz. Sie ändern ihre Position allmählich.
- *Appear*: Jeder Untertitel wird in Kopfrichtung vor dem Betrachter angezeigt (15° unter Augenhöhe). Er bleibt statisch in der Umgebung, auch wenn der Betrachter den Kopf bewegt. Die Position des nächsten Untertitels hängt von der neuen Kopfrichtung ab.

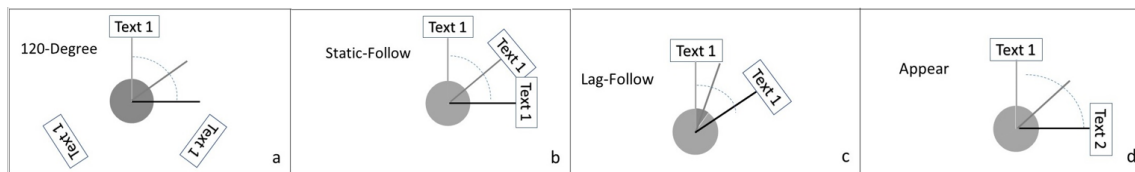


Abbildung 5-12: : Vier statische Untertitelmethoden. Die Untertitel: (a) sind dreimal immer am gleichen Ort platziert, (b) bewegen sich mit dem Kopf des Betrachters mit, (c) bewegen sich, wenn sich der Kopf um mehr als 30° dreht, (d) wenn sich der Untertitel ändert, wird er dort platziert, wo der Zuschauer hinschaut.

5.3.2 Herausforderungen für barrierefreie CVR-Experiences

Um CVR barrierefrei zu gestalten, gibt es mehrere wesentliche Herausforderungen:

Herausforderung 1 – Position von Untertiteln/Gebärdendolmetscher: In einem sphärischen Film gibt es keinen unteren Rand, so dass die Standardposition für Untertitel/Gebärdendolmetscher im Film nicht verfügbar ist. Für Untertitel in CVR sind sowohl Positionen im Film als auch auf dem Display möglich.

Herausforderung 2 – Sprecheridentifikation: Herauszufinden auf welche Person sich eine eingeblendete Zusatzinformation bezieht, ist in CVR relevanter als in framed Filmen, da alle Charaktere einer Szene im sphärischen Bild vorhanden sind, auch wenn der Betrachter nur einen Teil davon im FoV sieht. Die Platzierung der Untertitel/ Gebärdendolmetscher in der Nähe des Sprechers erleichtert die Identifizierung des Sprechers, schränkt jedoch die Blickrichtung des Betrachters beim Verfolgen des Textes ein.

Herausforderung 3 – Audiodeskription: Menschen mit Sehbehinderung benötigen zusätzliche Auskünfte zu den visuellen Informationen. Diese sollten von der gewählten Blickrichtung abhängen und nur den Bildausschnitt beschreiben, der im HMD angezeigt wird. Für sehbehinderte Menschen ist räumlicher Sound besonders wichtig, um eine Orientierung in VR zu ermöglichen.

Herausforderung 4 – Leichte Sprache: Auch Zusatzinformationen in leichter Sprache sollten die aktuelle Blickrichtung berücksichtigen und sich auf den Bildausschnitt beziehen, der im HMD angezeigt wird. Möglicherweise braucht es zusätzliche Hinweise, um auf Bereiche außerhalb des FoVs aufmerksam zu machen.

Herausforderung 5 – Rotation bei körperlichen Einschränkungen: Nicht jeder Zuschauer kann einer CVR-Anwendung ohne Weiteres durch Körper- oder Kopfdrehungen folgen. Hier sind Alternativen nötig, um CVR erlebbar zu machen.

5.3.3 Methoden für barrierefreie CVR-Experiences

Da die Barrierefreiheit in CVR ein sehr umfangreiches und komplexes Gebiet ist, haben wir uns in einem ersten Schritt auf Methoden für Menschen mit Hörbeeinträchtigungen fokussiert. Für andere Aspekte der Barrierefreiheit werden in Abschnitt 5.3.5 Lösungsansätze vorgeschlagen und diskutiert.

Um die Vor- und Nachteile von *screen*- und *world*-referenzierte Untertiteln zu analysieren, wurden drei Studien durchgeführt: (1) Vergleich von *screen*- und *world*-referenzierten Untertitel mit Namen zur Sprecheridentifizierung – hörende Teilnehmer, (2) Vergleich von *screen*- und *world*-referenzierten Untertiteln mit Farben zur Sprecheridentifizierung – Teilnehmer mit Höreinschränkungen, (3) dynamischer Wechsel zwischen *screen*- und *world*-referenzierten Untertiteln.

5.3.3.1 Studie: Untertitel für Fremdsprachen

Als erste Annäherung an dieses Thema wurde **Studie 7.4.5** mit hörenden Teilnehmern realisiert, um Untertitel-Methoden für fremdsprachige Filme zu finden. Dazu wurden zunächst verschiedene Positionen von *screen*-referenzierten Untertiteln getestet, die Methoden aus Abschnitt 5.3.1 implementiert

(Abbildung 5-12) und in einer Vorstudie miteinander verglichen. Dabei wurde die Static-Follow-Methode, von allen Benutzern unserer Vorstudie bevorzugt. Dieses Ergebnis stimmt mit der Studie der BBC-Research-Abteilung überein (Brown u. a., 2018), deren Ergebnisse damals noch nicht veröffentlicht waren. In einem zweiten Schritt wurde diese Methode in der Hauptstudie mit *world*-referenzierten Untertiteln, die in die Filmwelt integriert sind, verglichen:

- **Untertitel - *screen*-referenziert, mit Namen:** Unabhängig von der Blickrichtung sind die Untertitel am unteren Rand des Displays platziert. Zur Sprecheridentifikation ist der Name des Sprechers am Anfang des Textes hinzugefügt.
- **Untertitel - *world*-referenziert, mit Namen:** Der Text ist mit der Filmwelt verbunden und in der Nähe des Sprechers platziert. Die genaue Position hängt vom Filmszenario ab. Zur Sprecheridentifikation ist der Name des Sprechers am Anfang des Textes hinzugefügt.

Ergebnisse: Vergleich *screen*- und *world*-referenzierten Untertitel

Bei den Fragen zum direkten Vergleich, wurde keine generell bevorzugte Methode gefunden. *Screen*-referenzierte Untertitel erleichtern das Umhersehen, erschweren jedoch die Verfolgung des Inhaltes. Für die Sprecheridentifikation ist das *world*-referenzierte Verfahren geeigneter. Der Betrachter kann die sprechende Person sehen und die Untertitel gleichzeitig ohne extensive Augenbewegungen lesen. Auf diese Weise ist es einfacher, den Inhalt zu erfassen. Wenn jedoch die sprechende Person wechselt und die folgende Person nicht im FoV ist, ist ein gewisser Aufwand nötig, um den neuen Sprecher und die Untertitel zu finden. Enthält der Film mehrere Personen, ist es schwierig, die Untertitel mit der *screen*-referenzierten Methode den Sprechern zuzuordnen und es ist daher schwerer, den Film zu verstehen.

Auch wenn die Teilnehmer keine der Methoden bevorzugten, erzielten die Fragen zum VR-Erlebnis bessere Ergebnisse für die *world*-referenzierte Methode. Ein Grund könnte sein, dass *world*-referenzierte Untertitel in den Film integriert und *screen*-referenzierte Untertitel Teil des Displays sind. Die Teilnehmer merkten an, dass *world*-referenzierte Untertitel „natürlicher sind, sie stimmen eher mit dem wirklichen Leben überein“. In Bezug auf Task Workload, Simulator-Sickness und Presence schnitt in mehreren Fällen die *world*-referenzierte Untertitelung besser ab. Sie ist weniger anstrengend für die Augen, da die Untertitel in der Nähe der sprechenden Personen platziert sind und der Betrachter nicht gezwungen ist, in den unteren Bereich des Displays zu wechseln.

In unserem Experiment führte die *screen*-referenzierte Methode meist zu einer höheren Belastung. Eine Erklärung dafür könnte sein, dass das Lesen von Untertiteln und das gleichzeitige Ansehen des Films mehr Aufwand erfordert, wenn Sprecher und Text nicht verknüpft sind. Eine Ausnahme war die höhere körperliche Belastung der *world*-referenzierten Methode in einer Szene mit vielen Personen. Dies kann dadurch verursacht sein, dass die Personen dort nicht nahe beieinander waren, so dass der Betrachter die nächste Person zum Lesen des Untertitels suchen und auch den Kopf bewegen musste.

Während der Versuche trat kaum Simulator-Sickness auf. Bei den meisten Items war der Unterschied zwischen den Methoden gering, bei der *screen*-referenzierten Methode war der Score für die Augenbelastung jedoch höher, da mehr Augenbewegungen für den Wechsel zwischen Sprecher und Text erforderlich waren.

Wir sind uns der Tatsache bewusst, dass nicht alle Ergebnisse auf Untertitel für gehörlose Zuschauer übertragbar sind. Auch wenn bei framed Filmen die Untertitel für Hörende und Gehörlose meist gleich sind, macht es in virtuellen Umgebungen einen großen Unterschied, ob der Betrachter die sprechende Person hören kann, da es von der Blickrichtung abhängt, ob diese auch visuell wahrnehmbar ist. Eine hörende Person bemerkt, wenn eine Person zu sprechen beginnt, wogegen gehörlose Menschen auf visuelle Informationen angewiesen sind. Teile der **Studie 7.4.5** wurden zusätzlich mit einer gehörlosen Teilnehmerin durchgeführt, die uns wertvolle Hinweise für die weiteren Studien gab. Die Namen in den Untertiteln der **Studie 7.4.5** wurden auf dieser Grundlage in den **Studien 7.4.7** und **7.4.8** durch Farben, die die verschiedenen Sprecher anzeigen, ersetzt.

5.3.3.2 Studie: Untertitel für gehörlose Nutzer

Für die **Studie 7.4.7** mit gehörlosen Teilnehmern wurden zwei verschiedene Animationsfilme mit einer großen Anzahl an Charakteren ausgewählt, um eventuelle Probleme in der Sprecheridentifizierung herauszufinden. In einem Video gab es fünf sprechende Charaktere, im anderen zwei mit einer Reihe von nicht sprechenden Nebencharakteren. Es wurden folgende Methoden miteinander verglichen:

- **Untertitel – *screen*-referenziert, farblich:** Der Text ist statisch mit dem HMD verbunden und bewegt sich bei Kopfbewegungen mit. Unabhängig von der Blickrichtung sind die Untertitel am unteren Rand des Displays platziert. Die Texte haben für jeden Sprecher eine andere Farbe.
- **Untertitel – *world*-referenziert, farblich:** Der Text ist mit der Filmwelt verbunden und in der Nähe des Sprechers platziert. Die genaue Position hängt vom Filmszenario ab. Die Texte haben für jeden Sprecher eine andere Farbe.

Ergebnisse

In Studie 7.4.7 empfanden die Teilnehmer die *screen*-referenzierten Untertitel leichter lesbar als die *world*-referenzierten. Dieses schlechtere Abschneiden der *world*-referenzierten Untertitel im Gegensatz zu Studie 7.4.5 ist möglicherweise darauf zurückführbar, dass es für gehörlose Anwender schwierig ist ohne zusätzliche Hilfsmittel zu erkennen, ob eine Person spricht, die nicht im FoV ist. Die Teilnehmer wünschten sich einen Wechsel zwischen den beiden Methoden, und empfahlen zusätzlich einen Pfeil als Unterstützung der Sprecheridentifikation.

Aufgrund der Ergebnisse aus **Studien 7.4.5** und **7.4.7** wurde eine Kombination von *screen*- und *world*-referenzierten Untertiteln implementiert (**Studie 7.4.8**), bei der automatisch zwischen beiden Methoden hin- und hergewechselt wurde. Waren Sprecher und *world*-referenzierte Untertitel im FoV, wurden nur diese angezeigt. Bewegte der Zuschauer seinen Kopf so, dass diese nicht mehr sichtbar waren, wechselte der Untertiteltyp zu *screen*-referenziert, und ein Pfeil gab zusätzlich an, in welcher Richtung sich der Sprecher befindet. Als Alternative wurde auch noch die in 7.3.4 verwendete Linie als Guiding Methode getestet, welche aber zu sehr ablenkte. Der Pfeil erwies sich als hilfreich und verbesserte die Sprecheridentifikation.

5.3.3.3 Studie: Gebärdendolmetscher

Da viele Menschen mit Hörbehinderungen Gebärdendolmetscher den Untertiteln vorziehen, wurden auch diese in unsere Studien aufgenommen:

- **Gebärdendolmetscher *border*-Methode:** Das Bild des Gebärdendolmetschers ist farblich umrandet. Jeder sprechenden Person wird dabei eine andere Farbe zugeordnet.
- **Gebärdendolmetscher *image*-Methode:** Dem Gebärdendolmetscher ist ein Bild des Sprechers beigefügt.

Ergebnisse

In **Studie 7.4.6** wurden zwei Methoden für die Integration des Gebärdendolmetschers miteinander verglichen: *image*- und *border*-Methode. Die *image*-Methode war dabei hilfreicher als die *border*-Methode. Jedoch hatten Teilnehmer in beiden Methoden Schwierigkeiten, dem Gebärdendolmetscher und dem Video gleichzeitig zu folgen. Da die Randbereiche auf HTC-Vive-Displays momentan nur unscharfe Darstellungen ermöglichen, ist es schwer, eine geeignete Position des Gebärdendolmetschers zu finden, so dass der Inhalt des Films nicht überdeckt wird. Die Teilnehmer wünschten sich deshalb ausgeschnittene Gebärdendolmetscher. Weiter Untersuchungen sind notwendig, um Methoden zu finden, wie Gebärdendolmetscher in sphärischen Videos mit mehreren Sprechern eingebettet werden sollten.

5.3.4 Folgerungen für die Herausforderungen

Zusammenfassend wurden folgende Ergebnisse gefunden:

- Sowohl *screen*-referenzierte als auch *world*-referenzierte Untertitel sind hilfreich. Eine Kombination der beiden Methoden, in Abhängigkeit davon, ob der Sprecher im FoV sichtbar ist, wird den Anforderungen am besten gerecht.
- Um geeignete Techniken zur Darstellung von Gebärdendolmetschern in CVR zu finden, sind weitere Untersuchungen notwendig.
- Es werden beim Einsatz von Untertiteln und Gebärdendolmetschern zusätzliche Methoden zur Sprecheridentifikation benötigt, z.B. unterschiedliche Farben oder zusätzliche Icons.

Die Ergebnisse unserer Studien liefern Ansätze für die Herausforderungen (1) Position von Untertiteln/Gebärdendolmetscher und (2) Sprecheridentifikation.

Für die Anforderungen (3) Audiodeskription, (4) Leichte Sprache und (5) Rotation bei körperlichen Einschränkungen sind weitere Untersuchungen notwendig, für die wir im nächsten Abschnitt einige Konzepte vorschlagen, die sich aus den Erfahrungen unserer Studien ergeben haben.

5.3.5 Weiterführende Konzepte und Ausblick

Neben den in den vorangegangenen Abschnitten untersuchten Methoden zur Bereitstellung von Untertiteln und Gebärdendolmetschern sind weitere Konzepte erforderlich, um Barrierefreiheit in CVR zu unterstützen.

Audiodeskription (Visuelle Einschränkungen)

Für Menschen mit Sehbehinderungen ist es besonders wichtig, dass CVR-Anwendungen mit spatial Sound angeboten werden. Dadurch haben auch sie die Möglichkeit eines immersiven Erlebnisses. Um die visuellen Elemente zu beschreiben, sind Konzepte für eine blickrichtungsabhängige Audiodeskription notwendig. Nur die Elemente sollten beschrieben werden, die im aktuellen FoV enthalten sind. Es ist möglich, dass dafür räumlicher Sound benutzt werden kann, um dem Zuschauer zusätzliche Unterstützung zu geben. Es sollte allerdings untersucht werden, ob dies mit dem Ton der CVR-Anwendung in Konflikt steht. Im europäischen IMac-Projekt (Immersive Accessibility, 2018) wurden erste Erfahrungen mit CVR-Zuschauern gesammelt, die auf Audiodeskription angewiesen sind (Fidyka und Matamala, 2018).

Leichte Sprache (Mentale Einschränkungen)

Um Menschen, die leichte Sprache zum Verständnis eines Filmes benötigen, in CVR zu unterstützen, werden Konzepte gebraucht, die ähnlich der Audiodeskription blickrichtungsabhängig sind. Das Senden von Zeichen, ähnlich der *smiley*-Methode, könnte Menschen mit geistigen Behinderungen zusätzlich dabei helfen, die Geschichte zu verfolgen.

Rotationsunterstützung (Körperliche Einschränkungen)

Menschen mit körperlicher Beeinträchtigung, die sich nicht ohne Weiteres drehen können (z. B. Rollstuhlfahrer) benötigen die Möglichkeit, die Blickrichtung auf eine andere Weise zu wählen, z. B. mit der oben beschriebenen *controller*-Methode. Aber auch die Verstärkung der Kopffrotation kann für Betroffene eine Unterstützung bieten. Ein weiteres mögliches Konzept ist die Verwendung von Brain-Controlled User Interfaces (BCIs), um das Umherschauen in CVR ohne Körperbewegungen zu realisieren.

All diese Konzepte sind vielversprechende Ansätze, benötigen allerdings weitere Untersuchungen und eine Evaluierung ihrer Anwendbarkeit.

5.4 Realisierung von CVR-Player-Applikationen

5.4.1 Design-Space für CVR-Player-Anwendungen

Auf den Ergebnissen der Abschnitte 5.2 und 5.3 aufbauend wird der Design-Space aus Abschnitt 5.1.5 um zwei Unterdimensionen erweitert: den Gebrauch von Zusatzdiensten und der Rotation (in Tabelle 5-3 **fett** dargestellt).

	Dimension	Attribute (Optionen)			
Zuschauer	Distanz	<i>co-located</i>	remote		
	Anzahl	<i>Paar</i>	Gruppe (>2)		
	Rolle	<i>gleichrangig</i>	Guide	Follower	Slave
	Beziehung	<i>vertraut</i>	<i>anonym</i>		
	Gebrauch von Zusatzdiensten	Untertitel	Gebärdendolmetscher	Audiodeskription	Leichte Sprache
Nachricht	Kanal	<i>audio</i>	<i>visuell</i>	haptisch	Sensorinfo
	Trigger	<i>kontinuierlich</i>	<i>ausgelöst durch Sender</i>	deaktivierbar durch Empfänger	ausgelöst durch Sensoren
	Type	Information	Emotion	Richtung, Positionen	
	Referenz	screen-referenziert	world-referenziert	switching	
Gerät	Input	<i>Kopf/Blick</i>	Sprache	<i>Controller</i>	Gesten
	Rotation	Kopf-normal	Kopf-Gain	Button	Geste/EEG
	Display	<i>HMD</i>	Monitor	mobile Device	CAVE
	Symmetrie	<i>symmetrisch</i>	asymmetrisch		

Tabelle 5-3: Design-Space für CVR-Player-Applikationen. *kursiv*: in in den Studien dieser Arbeit implementierte Attribute. grau: von außen durch die teilnehmenden Personen, Räumlichkeiten und verwendeten Geräte gegeben. blau: vom Anwender konfigurierbar. **Fett**: neu hinzugekommenen Dimensionen.

Die im Design-Space aufgezeigten Dimensionen helfen nicht nur den CVR-Entwicklern alle wichtigen Anwender-Aspekte einer CVR-Applikation zu beachten. Sie sind auch hilfreich, um ein allgemeines Konzept für einen CVR-Player zu entwickeln. Die einzelnen Dimensionen können dafür in zwei Gruppen kategorisiert werden: solche, die vom Nutzer konfigurierbar sind (in Tabelle 5-3 blau unterlegt) und solche, die durch die teilnehmenden Personen, Räumlichkeiten und verwendeten Geräte vorgegeben sind (in Tabelle 5-3 grau unterlegt). Bei der Konzipierung eines CVR-Players müssen die vom Nutzer konfigurierbaren Attribute berücksichtigt werden.

5.4.2 Konzept für einen CVR-Player

Für alle gängigen VR-Plattformen gibt es Movie-Player zum Anzeigen linearer, timeline-basierter CVR-Applikationen (360°-Player). Nichtlineare, interaktive CVR-Applikationen werden momentan genau wie andere VR-Applikationen als Anwendung für die jeweilige VR-Plattform realisiert, z. B. als App für Oculus, Google Daydream, Vive oder auch Android-Apps für Cardboards. Dabei ist jede dieser Applikationen eine individuell programmierte Anwendung. Mit dem Spaceline-Konzept in Kapitel 4, dem dazugehörigen CVR-Editor und dem CVR-Containerformat wird ein erster Schritt unternommen, auch Filmemachern ohne Programmiererfahrung die Realisierung nichtlinearer, interaktiver CVR-Projekte zu ermöglichen. Damit die mit dem CVR-Editor erzeugten nichtlinearen, interaktiven CVR-Applikationen angeschaut werden können, ist eine Unterstützung durch ein Playerprogramm erforderlich, welches die Metadaten interpretiert und die Mediendaten entsprechend ausgibt. Aus den Erfahrungen der Studien in diesem Kapitel, dem Design-Space und den Erkenntnissen der Literaturrecherche wird im folgenden Abschnitt ein Konzept für einen CVR-Media-Player und dem dazugehörigen Containerformat vorgestellt.

Aufgabe des CVR-Media-Players ist es, dem CVR-Nutzer ein optimales CVR-Erlebnis zu ermöglichen. Dabei sollen Komfort und Barrierefreiheit berücksichtigt werden, das Spaceline-Konzept umgesetzt werden und auch Kontakte zwischen CVR-Nutzern möglich sein. Aus den bisherigen Erkenntnissen ergeben sich die folgenden Anforderungen an einen CVR-Media-Player:

Komfort:

- Möglichkeit die Kopfdrehung durch Rotation Gain oder Controller zu unterstützen oder zu ersetzen
- Unterstützung verschiedener Bereichs-Formate (z. B. 180°, 225°, 360°)

Barrierefreiheit:

- Gebärdendolmetscher als screen-referenziertes PiP
- Untertitel dynamisch wechselnd zwischen world- und screen-referenziert
- Leichte Sprache als zusätzliche Audiospur, Inhalt abhängig von der Kopfrichtung
- Audiodeskription als zusätzliche Audiospur, Inhalt abhängig von der Kopfrichtung
- Möglichkeit die Kopfdrehung durch Rotation Gain oder Controller zu unterstützen oder zu ersetzen

Spaceline:

- Interpretation und Umsetzung der Regionseigenschaften
- Interpretation und Umsetzung der Pointereigenschaften
- Interpretation und Umsetzung der Triggereigenschaften
- Interpretation und Umsetzung der Transitioneigenschaften

Soziale Aspekte:

- Unterstützung von Voice-Chat
- Unterstützung des Sendens von Rol-positionierten Smileys
- Unterstützung von Video-Chat

5.4.2.1 CVR-Containerformat

Für die Speicherung der für eine CVR-Applikation notwendigen Daten und für den Austausch zwischen den Programmen wird ein CVR-Containerformat benötigt. Neben den Bild- und Audiodaten jedes Spaces sind die Metadaten, die Beschreibung der Regionen, des Pointers, der Trigger-Ereignisse und Transitions notwendig. Basierend auf den Ergebnissen dieses Kapitels sind die folgenden Komponenten zum CVR-Containerformat aus Abschnitt 4.3.3.4 zu ergänzen: Untertitel, Gebärdensprache, Audiodeskription und leichte Sprache. Abbildung 5-13 illustriert die notwendigen Elemente eines CVR-Containerformates in einer Übersicht.

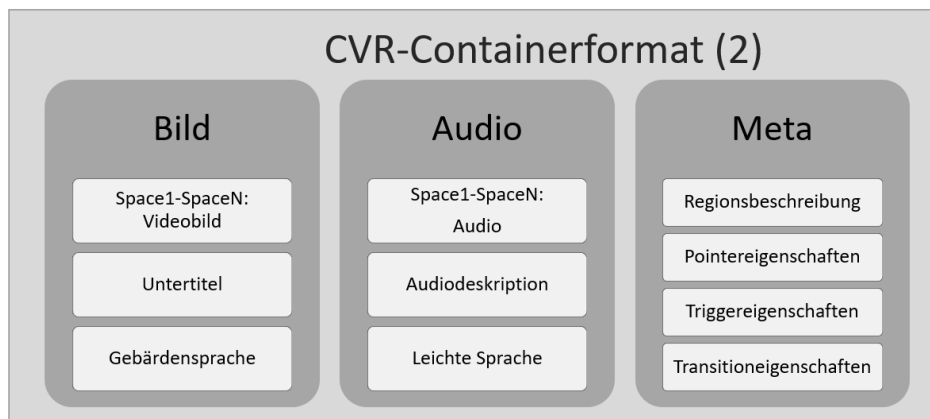


Abbildung 5-13: Notwendige Bestandteile eines CVR-Containerformates, um Spacelilne-Konstrukte und Elemente zur Barrierefreiheit zu speichern.

Im Jahre 2000 wurde von Lasse Kärkkäinen die Entwicklung des Multimedia Container Formates MCF gestartet, welches neben Video- und Audiostreamdateien auch Untertitel enthält. Es war das erste offene und flexible Mediencontainerformat, welches mehrere Streams dieser drei Typen zusammenfassen konnte. Das sich später daraus entwickelnde Multimedia-Containerformat Matroska (.mkv) (Matroska, 2018) ermöglicht die Einteilung in Kapitel, Verzweigungen und die Einbettung anderer Medienformate wie Bilder, pdf-Dateien und Texten. Eine Untermenge dieses Containerformates ist WebM (WebM-Consortium, 2019) für Web-Browser. Matroska ist ein offenes Standardprojekt und verwendet EBML (Extensible Binary Meta Language), zur Beschreibung des Containers. EBML ist ein binäres XML-Format zur Speicherung und Bearbeitung von Daten in hierarchischer Form (Matroska, 2015) und ermöglicht zukünftige Formaterweiterungen. Die technischen Spezifikationen, die den Bitstream beschreiben, sind für jeden offen (Matroska, 2020) und bieten eine Möglichkeit das CVR-Containerformat umzusetzen

5.4.2.2 Komponenten

Die Konzeption eines CVR-Movie-Players erfordert die Auseinandersetzung mit zahlreichen Aspekten. Da sich diese Arbeit auf die Mensch-Computer-Schnittstelle konzentriert, werden im Folgenden die Konfigurationsmöglichkeiten eines solchen Players näher betrachtet.

Um die obigen Anforderungen umzusetzen, sollte der CVR-Player so konfigurierbar sein, dass er an die Bedürfnisse des Nutzers angepasst werden kann und die Anforderungen der jeweiligen CVR-Anwendung berücksichtigt. Abbildung 5-14 illustriert die Komponenten eines CVR-Movie-Players in einer Übersicht.

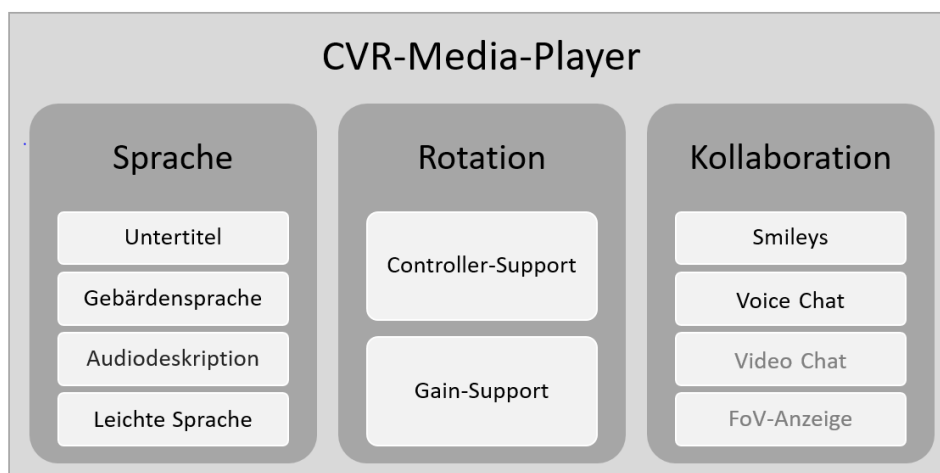


Abbildung 5-14: Komponenten eines CVR-Movieplayers, um Spacelilne-Konstrukte, Barrierefreiheit und gemeinsame CVR-Erlebnisse zu ermöglichen.

Für eine übersichtliche Darstellung der Struktur des Menüs eines CVR-Players wird in Abbildung 5-15 die UML-Syntax genutzt und aufgezeigt, welche notwendigen Parameter vom Nutzer eingestellt werden können.

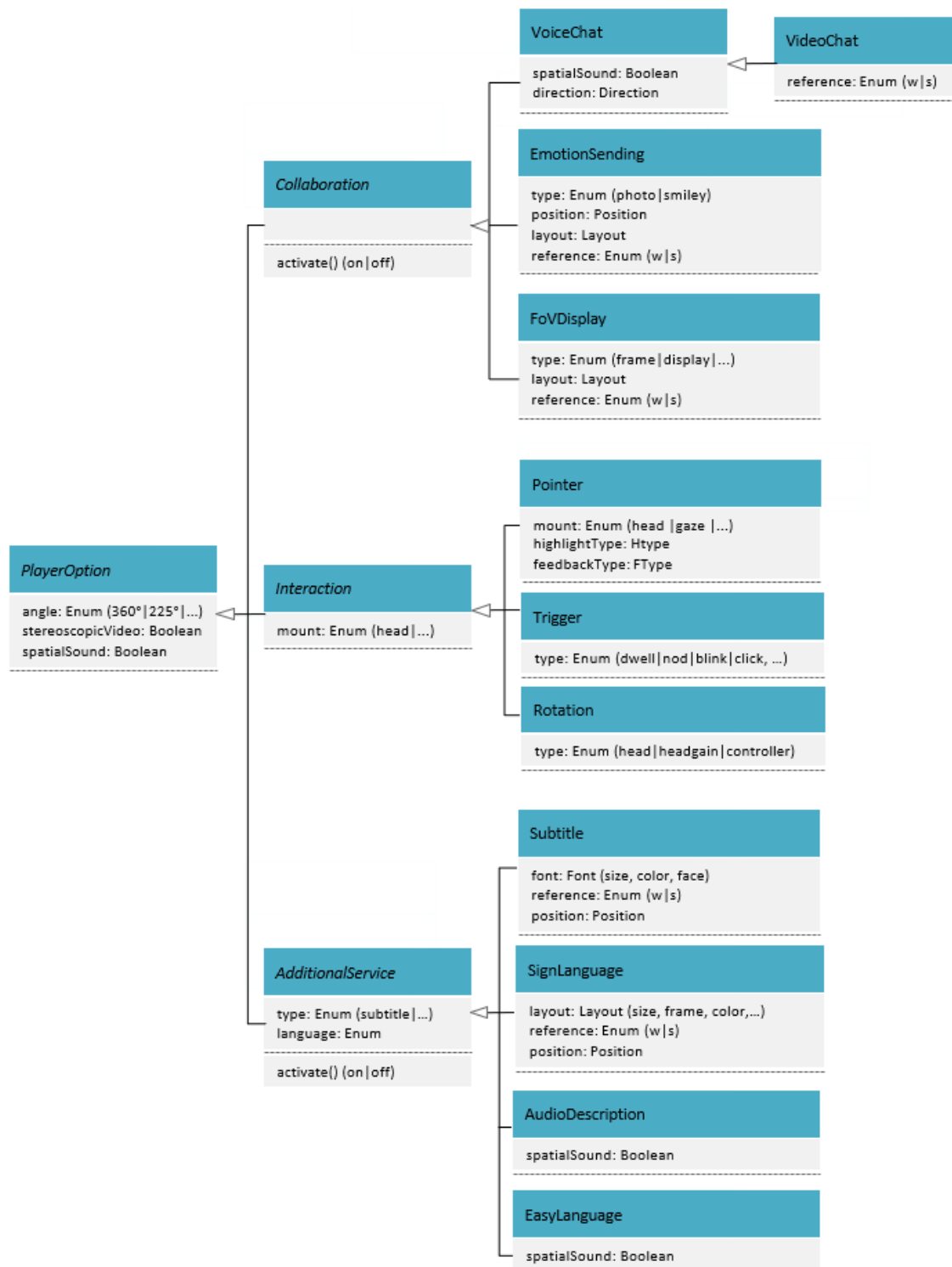


Abbildung 5-15: Die vom Nutzer konfigurierbaren Parameter des CVR-Players.

5.5 Fazit und Ausblick

Die Ergebnisse dieses Kapitels tragen dazu bei, Wege zu finden, wie CVR komfortabler und barrierefrei gestaltet werden kann und unterstützen den sozialen Aspekt von CVR-Applikationen.

In diesem Kapitel wurden Herausforderungen beschrieben, die für die Nutzung von CVR relevant sind, Dabei wurden soziale Aspekte ebenso berücksichtigt, wie Nutzerkomfort und Barrierefreiheit. Mehrere Ansätze wurden vorgestellt, implementiert, miteinander verglichen und diskutiert. Alle Ergebnisse hängen jedoch vom Filminhalt und der Testumgebung ab und weitere Studien für unterschiedliche Filmgenres, verschiedenen Geräten und anderen Teilnehmergruppen (insbesondere Teilnehmer, die Zusatzdienste benötigen) sind erforderlich, um allgemeingültige Aussagen zu erhalten.

Basierend auf dem gesammelten Wissen wurde ein Design-Space für CVR-Player-Applikationen mit drei Hauptdimensionen vorgestellt: Zuschauer, Benachrichtigung und Gerät. Wir haben diesen Design-Space auf der Basis der untersuchten Ansätze diskutiert und zusätzliche Aspekte für zukünftige Forschungen aufgezeigt. Der eingeführte Design-Space unterstützt die weitere Entwicklung von CVR-Applikationen. Er bietet einen strukturierten Ansatz für die Entwicklung neuer Ideen und Konzepte. Ein erster Anwendungsfall, ein Konzept für einen CVR-Player, wurde in diesem Kapitel umgesetzt.

6

How to Evaluate: Analysemethoden für CVR

„Noch nie ist eine Kunst groß geworden ohne Theorie. ... Ist es denn nicht die Theorie, die den Dingen erst die Würde verleiht Der Film ist eine Tatsache, eine so allgemeine sozial und psychisch so tiefwirkende Tatsache geworden, daß wir gerne oder nicht, uns mit ihr auseinandersetzen müssen.“

Béla Balázs (Balázs, 1924)

Um das Sehverhalten der Benutzer in CVR zu untersuchen, sind Methoden zum Erheben und Analysieren von Daten erforderlich. Wir haben dafür ein Analysetool, den CVR-Analyzer, entwickelt, mit dem Head- und Eyetracking-Daten von CVR-Zuschauern analysiert werden können. Die Daten werden auf einer äquirektangulären Projektion des Films als dynamische Overlays von Tracks und Heatmaps synchron zum Timecode des Films visualisiert und ermöglichen es, das Sehverhalten der Benutzer explorativ zu untersuchen und zu vergleichen. Des Weiteren werden in diesem Kapitel Methoden der räumlichen Statistik, die für Signifikanzaussagen genutzt werden, erläutert und etablierte Fragebögen, die für CVR geeignet sind und in den Studien verwendet wurden, vorgestellt.

Dieses Kapitel enthält bereits veröffentlichtes Material aus folgenden Publikationen:

- Rothe, S., Höllerer, T., Hußmann, H., 2018. CVR-Analyzer: A Tool for Analyzing Cinematic Virtual Reality Viewing Patterns. in: Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia - MUM '18, 127–137
- Rothe, S., Hussmann, H., 2018. Spatial statistics for analyzing data in cinematic virtual reality. in: International Conference on Advanced Visual Interfaces - AVI '18, Grosseto, Italy

Bei der Umsetzung der Ideen und der Implementierung des CVR-Analyzers wurde ich von Detlef Rothe unterstützt.

In CVR-Studien ist es notwendig, unterschiedliche Typen von Daten auszuwerten:

- Tracking-Daten, die Kopf- und Blickrichtungen beschreiben
- Antworten auf Fragebögen
- Interviewaussagen
- Andere gemessene Werte, wie Zeiten o.ä.

Für die getrackten Kopf- und Augenbewegungen, ist es wichtig, dass diese gemeinsam mit dem dazugehörigen Film analysiert werden. Genau wie bei anderen Daten sind zum Analysieren von Tracking-Daten unterschiedliche statistische Methoden nutzbar, welche in deskriptive und induktive Methoden unterteilt werden.

Für die deskriptive Darstellung der Daten wurde von uns ein Tool, der CVR-Analyzer, entwickelt, der in Abschnitt 6.2 vorgestellt wird. Für die Untersuchung von Signifikanzen sind Methoden der räumlichen Statistik anwendbar, die in Abschnitt 6.3 beschrieben werden.

Für die Analyse der anderen drei Datentypen sind die in der HCI-Community üblichen statistischen Methoden anwendbar.

Um Faktoren wie Presence oder Simulator-Sickness in VR messen zu können, existieren etablierte Fragebögen, die in Abschnitt 6.4 diskutiert werden.

6.1 Theoretischer Hintergrund

Deskriptive Statistik für Head- und Eyetracking-Daten

Für die deskriptive Analyse von Head- und Eyetracking-Daten sind unterschiedliche Methoden anwendbar. Pausch et al. (Pausch u. a., 1996) nutzten *Histogramme*, aus denen hervorgeht, wie weit die Zuschauer ihren Kopf horizontal und vertikal bewegen. Andere Studien ermittelten den Prozentsatz der Personen, die die Rols im Laufe der Zeit gesehen hatten (Sheikh u. a., 2016) oder visualisierten die Daten in Diagrammen und Tabellen (Nielsen u. a., 2016; Sheikh u. a., 2016).

Es gibt mehrere Arbeiten über Datensets von Kopf- und Augenbewegungen beim Betrachten omnidirektionaler Bilder (De Abreu u. a., 2017; Li u. a., 2017; Rai u. a., 2017; Upenik und Ebrahimi, 2017). Die erhaltenen Ergebnisse über Saliency Maps sind nützlich für Rendering-, Übertragungs- und Komprimierungsalgorithmen (De Abreu u. a., 2017; Rai u. a., 2017). Upenik und Ebrahimi (Upenic und Ebrahimi, 2017) beschrieben eine Methode zum Ermitteln von kontinuierlichen fixation-Maps. In unseren Studien variieren die Bilder im Laufe der Zeit kontinuierlich, so dass die zusätzliche Zeitkomponente bei der Analyse berücksichtigt werden muss.

Corbillon et al. (Corbillon u. a., 2017) untersuchten Datensätze von Kopfbewegungen und präsentierten die Ergebnisse in Diagrammen und Tabellen. Ein weiterer Ansatz zur Visualisierung von Raum-Zeit Daten sind Raum-Zeit-Würfel (Bach u. a., 2014; Drusch u. a., 2014), die aus den zwei Raumkoordinaten und der Zeitkoordinate gebildet werden. Da sich unsere Raumdaten auf einer Kugel befinden, müssen sie dazu auf

die euklidische Ebene projiziert werden. Dadurch sind die Bereiche am linken und rechten Rand getrennt, auch wenn sie im Original nahe beieinander liegen. Darüber hinaus ist ein gewisser Aufwand erforderlich, um die Zeitscheiben des Würfels mit den Bildern im Film abzugleichen.

Es gibt verschiedene Methoden zur Visualisierung von Eyetracking-Daten, die auch für Headtracking-Daten angewendet werden können. Blascheck et al. (Blascheck u. a., 2017) klassifizierten solche Methoden und analysierten 110 Studien, die Visualisierungstechniken für Eyetracking verwendeten. Dabei wurden unterschiedliche Arten von Reizen untersucht. Videos sind dynamische Reize, die nicht nur eine Raumkomponente, sondern auch eine Zeitkomponente haben. Mit bewegungskompensierten Maps (Kurzhaus und Weiskopf, 2013) können bewegte Objekte beobachtet werden. Das Ergebnis ist jedoch eine statische Map ohne Visualisierung der zeitlichen Blickinformationen. Burch et al. (Burch u. a., 2013) führten die AOI-Flusstechnik zur Visualisierung zeitlich variierender Frequenzen von Blickfixierungen ein und zeigten die Daten als flussähnliche Strukturen. Für unsere Forschung ist es jedoch wichtig, die Daten direkt auf dem Film zu visualisieren.

Auch einige kommerzielle Produkte (Unity, 2017; Viar360, 2016; Wistia, 2017) ermöglichen die Analyse des Nutzerverhaltens. Die Video-Heatmaps von Wistia (Wistia, 2017) bieten kontextbezogene Informationen (wann und wo das Video angesehen wurde). In einer farbcodierten Timeline wird dargestellt, wie der Betrachter mit dem Video interagierte, welche Teile des Videos übersprungen und welche Teile wie oft gesehen wurden. Diese farbliche Kennzeichnung wird auch Heatmap genannt. Uns interessiert dagegen die Blickrichtung des Nutzers, die wir auf den Filmbildern visualisieren. Viar360 (Viar360, 2016) bietet Heatmaps für Filmbilder. Die visualisierten Daten zeigen jedoch nicht die Blickrichtung. Produzenten können Pols definieren und die Heatmap zeigt, wie oft Benutzer diese aktivieren. Da sphärische Videos auch für Werbezwecke verwendet werden, interessieren sich die Produzenten für die Informationen, nach denen der Benutzer sucht. Daher entstehen immer mehr solcher Plattformen, die wir nicht alle inspizieren konnten. Einige Tools benötigen eine Mindestanzahl von Betrachtungen zum Generieren von Heatmaps: YouTube-Heatmaps (YouTube, 2017) mindestens 1000, Facebook (facebook, 2016) mehr als 50.000 Zuschauer. In unseren Studien haben wir allerdings weniger Teilnehmer. In solchen Fällen ist es notwendig, verschiedene Parameter bei der Generierung von Heatmaps an die Studiensituation anzupassen. Alle diese Produkte setzen voraus, dass die Videos und Daten auf die Plattform des Anbieters gehostet werden.

Um die Tracking-Daten auf herkömmlichen Bildschirmen analysieren zu können, wird der CVR-Film in die Ebene projiziert. Dafür gibt es unterschiedliche Optionen, jede von ihnen verzerrt den Film auf unterschiedliche Weise. Map-Guides (De Abreu u. a., 2017; Snyder, 1987) können dabei helfen, die passendste Projektion zu finden. Jiang et al. (Jiang und Li, 2014) klassifizierten Map-Typen und untersuchten räumliche Interpolationsmethoden für verschiedene Arten von Projektionen. Wir haben verschiedene Projektionen geprüft und die Vor- und Nachteile abgewogen. Die äquirektanguläre Projektion der zweidimensionalen Kugel auf eine euklidische Ebene wurde gewählt, da sphärische Videos normalerweise in dieser Weise formatiert sind, und die anderen Projektionen keinen zusätzlichen Vorteil boten.

Mit unserem Ansatz möchten wir das Blickverhalten veranschaulichen, indem wir aufgezeichnete Daten als Überlagerung des Films visualisieren. Dies soll auf zwei Arten möglich sein: zum einen durch die Visualisierung der Tracks der Zuschauer, zum anderen durch Heatmaps die auf den Daten mehrerer Teilnehmer beruhen. In diesen Heatmaps werden Regionen, die besonders häufig angesehen werden, farblich hervorgehoben. Dadurch ist es möglich zu untersuchen, welche Cues die Aufmerksamkeit der Betrachter auf sich ziehen.

Induktive Statistik für Head- und Eyetracking-Daten

Genau wie bei anderen Datentypen ist es auch für Raum-Zeit-Datenanalysen notwendig zu wissen, ob die gewonnenen Ergebnisse signifikant sind. Die räumliche Statistik bietet dafür verschiedenen Methoden. Oft werden diese in geographischen und demographischen Forschungen angewandt. Dort haben die Daten oft eine ähnliche Struktur wie die in CVR getrackten Daten: Raumkoordinaten auf einer Kugeloberfläche und eine Zeitkoordinate.

Bei der Untersuchung des Sehverhaltens der CVR-Zuschauer ist es wichtig, zu identifizieren, in welche Bereiche besonders häufig oder selten geschaut wird. Bereiche, die besonders viele Blicke auf sich ziehen, *Hotspots* genannt, können durch Aufmerksamkeitslenkung entstehen. Der Begriff Hotspot wird in der Literatur unterschiedlich verwendet. In unserer Arbeit ist ein Hotspot ein Bereich mit einer hohen Anzahl von Tracking Punkten in einem bestimmten Raum-Zeit-Segment. In den Studien in Kapitel 7 werden solche Hotspots identifiziert und Hotspot-Bereiche für verschiedene Methoden miteinander verglichen.

Um zu ermitteln, ob die *Hotspots* signifikant oder eher durch zufällige Verteilungen entstanden sind, wurde von Getis und Ord die *Getis-Ord Gi** Statistik eingeführt (Getis und Ord, 1992; Ord und Getis, 1995). Diese identifiziert Regionen, in denen sich eine hohe oder niedrige Anzahl von Ereignissen sammelt.

6.2 CVR-Analyzer: Ein Tool zum Analysieren von Cinematic VR-Tracking-Daten

Für die Analyse unserer Studien war es notwendig, das Blickverhalten der Teilnehmer aufzuzeichnen und zu analysieren. Wie im vorherigen Abschnitt beschrieben sind Heatmaps sowie Track-Visualisierungen zwar etablierte Datenvisualisierungskonzepte für die Betrachtung von Bildern, die vorhandenen Tools genügten allerdings nicht den Anforderungen (Abschnitt 6.2.1), die sich aus unserer Arbeit ergaben. Aus diesem Grund wurde ein eigenes Analysetool für sphärische Filme entwickelt: der CVR-Analyzer.

Ähnlich wie bei geographischen Daten haben CVR Head- und Eyetracking-Daten zwei Koordinaten auf der Oberfläche einer Kugel, mit einem Azimut zwischen -180° und $+180^\circ$ und einer Elevation zwischen -90° und $+90^\circ$. Zur Visualisierung sphärischer Daten können diese auf eine zweidimensionale euklidische Ebene projiziert werden (Abbildung 6-1). Dies ändert jedoch die Metrik und führt zu Verzerrungen.

Sphärische Videos werden normalerweise in einer äquirektangulären Projektion der zweidimensionalen Kugel auf eine euklidische Ebene formatiert:

$$P: \begin{pmatrix} \lambda \\ \phi \end{pmatrix} \rightarrow \begin{pmatrix} x \\ y \end{pmatrix}, \quad \lambda, x \in [-\pi, +\pi], \quad \phi, y \in \left[-\frac{\pi}{2}, +\frac{\pi}{2}\right]$$

Die Betrachtungsrichtung ist ein Vektor (λ, ϕ) , wobei λ den Azimut (pitch) und ϕ die Elevation (yaw) darstellt (Abbildung 6-1).

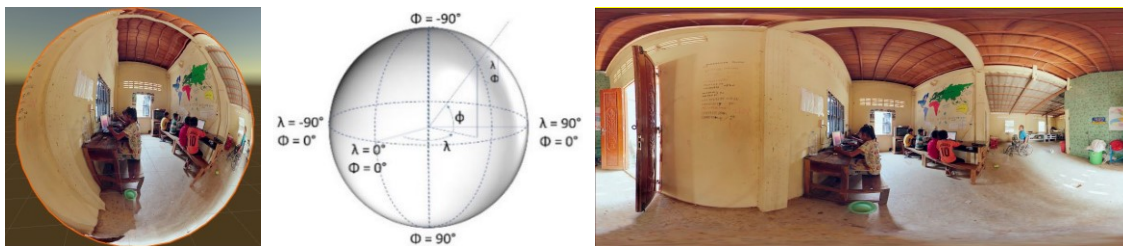


Abbildung 6-1: Die Koordinaten der Kugel für einen sphärischen Film (links), die äquirektanguläre Projektion (rechts).

Zur Visualisierung sphärischer Daten werden diese in gleicher Weise auf die zweidimensionale euklidische Ebene projiziert.

6.2.1 Anforderungen

Bei der Arbeit an unseren Studien wurden verschiedene Anforderungen für die Analyse von Raum-Zeit CVR-Daten identifiziert:

Anforderung 1: Verfügbarkeit und Performance

Das Tool soll ohne Installation leicht zugänglich sein und für eine Vielzahl von Zwecken eingesetzt werden können. Die Technologie der Datenaufzeichnung soll von der Datenanalyse getrennt sein. Zur Performanceverbesserung sollen Anpassungen der Aktualisierungsrate und anderer Parameter möglich sein.

Anforderung 2: Modi

Es wird ein Modus benötigt, der anzeigt, wo sich die Bereiche befinden, die von den Zuschauern am häufigsten angeschaut wurden (Heatmap-Modus). Zusätzlich soll es möglich sein, den Track jedes Teilnehmers oder einer Gruppe von Teilnehmern zu untersuchen (Track-Modus).

Anforderung 3: Projektion und Verzerrung

Bei der Projektion von Film und Daten in die Ebene ist zu berücksichtigen, dass die randlose Sphäre auf ein Rechteck projiziert wird und somit Punkte an den gegenüberliegenden seitlichen Rändern benachbarte Punkte auf der Sphäre sind. Um zu vermeiden, dass relevante Bereiche, die untersucht werden, an den Rändern liegen, sollte es möglich sein, die Ansicht zu drehen. Außerdem ändert die Projektion die Metrik und damit die Abstände zwischen Punkten. Eine richtige Interpretation sollte visuell unterstützt werden.

Anforderung 4: Zeitabhängigkeit

Neben der Raumkomponente gibt es in CVR die Zeitkomponente. Es muss berücksichtigt werden, dass sich das Bild zeitlich ändert. Im Gegensatz zu statischen Bildern, bei denen Nachbarpunkte in einem zweidimensionalen Raum analysiert werden, haben CVR-Daten drei Dimensionen.

Anforderung 5: Übersicht über Hotspots in der Timeline

Das Durchsuchen eines Films nach Hotspots ist sehr zeitaufwendig. Das Tool soll eine Übersichtsmöglichkeit bieten, Hotspot-Bereiche in der Timeline auf einen Blick zu erkennen.

Anforderung 6: Vergleich von Daten

Raum-Zeit Daten in CVR können für verschiedene Quellen aufgezeichnet werden: Headtracking-Daten, Eyetracking-Daten, Tracks von Stimuli (z. B. Flicker-Cues). In unseren Untersuchungen werden sowohl das gemeinsame wie auch separate Analysieren dieser Tracks benötigt. Darüber hinaus werden Vergleichsmöglichkeiten zwischen bestimmten Gruppen von Daten benötigt, wie z. B. Headtracking-Daten versus Eyetracking-Daten oder Vergleiche von Tracking-Daten, die für verschiedene Lenkungsmethoden in einer Studie aufgezeichnet wurden.

6.2.2 Implementierung

Aufzeichnung

Während der Nutzer einen CVR-Film ansieht, protokolliert ein Skript die Kopf- und / oder Augenbewegungen und den Timecode des Films in einem lesbaren, sehr einfachen Format. Die aufgezeichnete Datei enthält den Namen des Films, ein Identifizierungskennzeichen und die Koordinaten der einzelnen Tracking-Punkte (Abbildung 6-2).

```
movie: movieName
tag: E_20171102143744
time: 2017-11-02_14:37:44

# seekPos x y
0.002 6.397264 275.2098
0.017 6.357021 274.9597
0.034 8.262152 275.1749
...
152.254 22.63615 67.400
152.265 22.76232 67.438
```

Abbildung 6-2: Aufbau der Logdatei. Die Datei enthält am Anfang einige Metainformationen. Jeder Aufzeichnungseintrag umfasst drei Parameter: Den Timecode, die Elevation φ , die zwischen -90 und $+90$ variiert und das Azimut λ , welches zwischen 0 und $+360$ variiert.

Für unsere Untersuchungen wurde ein C#-Skript implementiert und Unity zum Generieren der Logdateien verwendet. Während der Wiedergabe eines CVR-Films protokolliert das Skript standardmäßig 60-mal pro Sekunde die Kopf- und/oder Augenbewegungen sowie den Timecode des Films. Die Aufzeichnungsfrequenz kann angepasst werden und ist eine Abwägung zwischen Performance und Genauigkeit. Der CVR-Analyzer arbeitet jedoch unabhängig von einer bestimmten Aufzeichnungsimplementierung, die Protokolldatei muss lediglich das in Abbildung 6-2 gezeigte Format aufweisen.

Verfügbarkeit (Anforderung 1)

Für einen einfachen Zugriff, ohne eine notwendige Installation, wurde das Tool mit den HTML5-Elementen *video* und *canvas* in HTML5 implementiert. Der CVR-Analyzer kann lokal oder remote über HTTP verwendet werden und erfordert einen Webserver (Abbildung 6-3). Zur Performanceoptimierung können der sphärische Film und die aufgezeichneten Daten auf dem Client bereitgestellt werden.

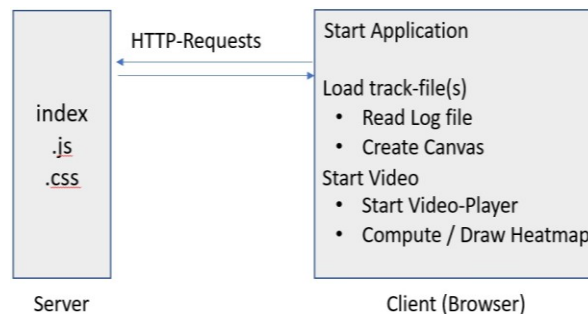


Abbildung 6-3: Grundlegende Softwarearchitektur und Datenfluss.

Beim Starten des CVR-Analyzers werden die Daten der ausgewählten Logfiles in den Speicher gelesen. Während der Wiedergabe des projizierten Films wird die JavaScript Methode `requestAnimationFrame()` rekursiv aufgerufen. Die Häufigkeit dieser Aufrufe ist hardwareabhängig (auf unserer Hardware ca. 60 Mal pro Sekunde). Jeder Aufruf verarbeitet die Daten, die in den Zeitraum seit dem letzten Aufruf fallen, und rendert den betroffenen Teil des Canvas.

Heatmap- und Track-Modus (Anforderung 2)

Zur Analyse kann zwischen zwei Modi gewählt werden und je nach gewähltem Modus werden Heatmaps oder Tracks berechnet und visualisiert. Die Heatmap illustriert, wie oft Zuschauer innerhalb eines bestimmten Zeitrahmens einen bestimmten Bereich betrachteten (Abbildung 6-4).



Abbildung 6-4: Ansicht des Films im Heatmap-Modus.

Da die Zuschauer ihre Aufmerksamkeit nicht genau auf den exakt gleichen Punkt des Films richten, ist ein Glättungsfilter erforderlich. Dies wurde durch eine kleine Zelle um den Betrachtungspunkt und einen radialen Gradienten realisiert. Für einen einzelnen isolierten Betrachtungspunkt gibt der Heatvalue die Anzahl der Tracking-Punkte in der umgebenden Zelle, gewichtet mittels der Gradientenfunktion, an. Auf ähnliche Weise wird die Opazität eines Punktes der Zelle berechnet. Eine einfache, von uns verwendete, Gradientenfunktion ist ein monoton abnehmender linearer Gradient.

$$grad(r) = S \begin{cases} \left(1 - \frac{1}{R} r\right) & r < R \\ 0 & r > R \end{cases},$$

wobei S ein Skalierungswert für die Zuordnung des Heatvalues zu einem Farbpunkt und R der Zellenradius auf der Kugel ist. Für die Berechnung der Heatmap auf der euklidischen Ebene muss die Verzerrung der Projektion berücksichtigt werden (siehe Abschnitt Projektion und Verzerrung).

Für Punkte in der Schnittmenge von Zellen werden die jeweiligen Heat- bzw. Opazitätswerte addiert.

$$h_t(x, y) = h_{<t}(x, y) + grad(\sqrt{(x - x_c)^2 + (y - y_c)^2} S),$$

$$h_0(x, y) = 0, t > 0$$

dabei ist $h_t(x, y)$ der Heatvalue am Punkt (x, y) für die Zeit kleiner t und $h_{<t}(x, y)$ der Heatvalue im Punkt (x, y) , falls ein neuer Trackpunkt bei (x_c, y_c) berücksichtigt werden muss. Eine ähnliche Funktion wird genutzt, um die Deckkraft an einem Punkt zu berechnen.

Der darzustellende Farbwert eines Punktes ergibt sich durch Abbildung des Heatvalues auf einer Farbskala (Abbildung 6-5: Farbskala, ein Wert von 0 entspricht Blau, ein Wert von 50 entspricht Gelb, ein Wert von 100 entspricht Rot.).

$$[0, 100] \rightarrow [rgb(0, 0, 255), rgb(255, 0, 0)]$$



Abbildung 6-5: Farbskala, ein Wert von 0 entspricht Blau, ein Wert von 50 entspricht Gelb, ein Wert von 100 entspricht Rot.

Die statische Festlegung eines Skalierungswertes S wirft ein Problem auf: Wird ein Heatvalue erreicht, der dem Maximalwert der Farbskala entspricht, werden keine Farbveränderungen mehr sichtbar. Um dies zu vermeiden und die Unterschiede über diesen Wert hinaus anzuzeigen, bietet das Tool verschiedene Optionen:

- Die Verwendung eines nichtlinearen Gradienten verringert dieses Problem, da der maximale Farbwert nur in der Nähe des Zellzentrums erreicht wird.
- Eine andere Möglichkeit besteht darin, den statischen Skalierungswert S durch einen dynamischen (S') zu ersetzen, der vom aktuellen Heatvalue abhängt (Abbildung 6-6).

$$S' = \left(1 - \frac{h_{\leq t}(x,y)}{100}\right) S$$



Abbildung 6-6: Verschiedene Skalierungsmethoden: eine statische Skalierung (links) – über dem Wert 100 ändert sich die Farbe nicht mehr; dynamische Skalierung (rechts) – der maximale Farbwert kann nicht erreicht werden.

Ein zweites Problem: Der Heatvalue und damit die Farbe sind abhängig von der Anzahl der Betrachter. Daher kann es schwierig werden, Tests mit einer unterschiedlichen Anzahl von Zuschauern zu vergleichen. Der CVR-Analyzer ist in der Lage Wertänderungen über eine unterschiedliche Anzahl von Zuschauern zu normalisieren.

Für eine optimale Untersuchung der Heatmaps können zahlreiche weitere Parameter eingestellt werden, z. B. Zellradius, Skalierung der Farbänderung und Opazität, Update-Raten.

Zusätzlich zum Heatmap-Modus bietet das Tool den Track-Modus und ermöglicht damit, die Tracks jedes Teilnehmers separat zu untersuchen (Abbildung 6-7).



Abbildung 6-7: Ansicht des Films im Track-Modus. Für jeden Teilnehmer wird ein Track angezeigt, der deaktiviert werden kann.

Projektion und Verzerrung (Anforderung 3)

Die Verwendung der äquirektangulären Projektion für den Film und der aufgezeichneten Daten hat den Vorteil, dass das gesamte Bild auf einmal angezeigt wird. Die Projektion ändert jedoch die Metrik: Je weiter ein Bereich von der horizontalen $\phi = 0^\circ$ -Linie (*Äquatorlinie*) entfernt ist, desto stärker wird der Bereich verzerrt (Abbildung 6-8).

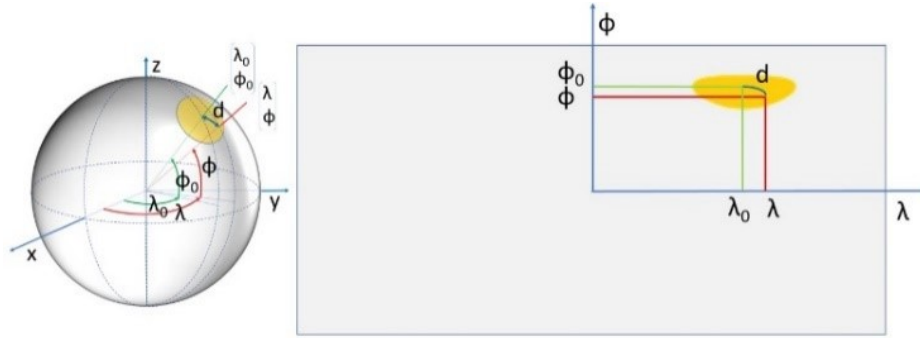


Abbildung 6-8: Die äquirektanguläre Projektion einer Kugel auf eine flache Ebene, die für den Film und die Daten verwendet wird.

Der Abstand d zwischen zwei Punkten (λ, ϕ) und (λ_0, ϕ_0) wird berechnet als:

$$d = \left\| \begin{pmatrix} \lambda \\ \phi \end{pmatrix} - \begin{pmatrix} \lambda_0 \\ \phi_0 \end{pmatrix} \right\|_{sphere} = \cos^{-1}(\sin \phi_0 \sin \phi + \cos \phi_0 \cos \phi \cos(\lambda_0 - \lambda))$$

Selbst wenn der Abstand zwischen zwei vertikalen Punkten (gleiches λ) auf der Kugel dem Abstand zwischen den projizierten Punkten entspricht,

$$\left\| \begin{pmatrix} \lambda \\ \phi \end{pmatrix} - \begin{pmatrix} \lambda \\ \phi_0 \end{pmatrix} \right\|_{sphere} = |\phi - \phi_0|$$

variiert der Abstand zwischen zwei horizontalen Punkten (gleiches ϕ).

$$\left\| \begin{pmatrix} \lambda \\ \phi \end{pmatrix} - \begin{pmatrix} \lambda_0 \\ \phi \end{pmatrix} \right\|_{sphere} = \cos^{-1}(\sin^2 \phi + \cos^2 \phi \cos(\lambda_0 - \lambda))$$

In der manuell erstellten Heatmap von Abbildung 6-9 haben alle angezeigten Bereiche auf der Kugel die gleiche Größe. Bei der Betrachtung solcher Heatmaps kann eine Fläche in Pol-Nähe ($\phi = \pm 90^\circ$) aufgrund ihrer Größe überschätzt werden. Deshalb kann der CVR-Analyzer-Benutzer die Verzerrung ausschalten. In diesem Fall werden die Heatmaps mit euklidischen Koordinaten berechnet, was wiederum zu einer Unterschätzung in den Polbereichen führen kann, da der Abstand zwischen diesen Punkten größer ist als auf der Kugel. In unseren eigenen Studien haben wir jedoch häufig von einem Wechsel zwischen beiden Modi profitiert. Außerdem befinden sich die meisten Head- und Eyetracking-Daten in CVR nahe der Äquatorlinie.



Abbildung 6-9: Eine manuell erstellte Heatmap zum Verständnis der Verzerrung im CVR-Analyzer in der Nähe der Pole.

Obwohl die äquirektanguläre Projektion (Abbildung 6-12) für CVR-Filme sehr gebräuchlich ist und es in HTML5 bereits einen Player dafür gibt, haben wir auch einige andere Projektionen mittels der Tissots Indikatrix (Clarke, 2017) untersucht (Mulcahy und Clarke, 2001) (Abbildung 6-10 – Abbildung 6-12). Eine Indikatrix beschreibt die Verzerrung an einem bestimmten Punkt. Bei anderen Projektionen, z. B. der *Hammerprojektion* (Abbildung 6-10), ist die Verzerrung an den Polen erheblich weniger ausgeprägt. Allerdings gibt es an den Rändern auch in der Nähe der Äquatorlinie eine größere Verzerrung. In unserer Forschung liegen die relevantesten Daten aber gerade in diesem Bereich. Andere Projektionen, wie die *azimutale äquidistante Projektion* (Abbildung 6-11), erscheinen aufgrund der geteilten Darstellung nicht geeignet, auch wenn die Verzerrungen geringer sind.

Die äquirektanguläre Projektion hat den Vorteil, Bereiche nahe der Äquatorlinie mit nur geringer Verzerrung darzustellen. Die Filmwiedergabe für diese Projektionen erfordert zudem weniger Rechenaufwand, da die Filme standardmäßig in diesem Format vorliegen. Da in CVR die getrackten Daten nur selten in Polnähe liegen, wurde die äquirektanguläre Projektion für den CVR-Analyzer gewählt.



Abbildung 6-10: Hammerprojektion mit Überlagerungen gemäß Tissotscher Indikatrix. Sie hat den Vorteil geringerer Verzerrung in den Polbereichen, aber den Nachteil einer stärkeren Verzerrung an den Rändern.



Abbildung 6-11: Azimutale äquidistante Projektion (Hemisphären) mit Überlagerungen gemäß Tissotscher Indikatrix. Sie hat den Vorteil einer geringeren Verzerrung, aber den Nachteil einer geteilten Anzeige.

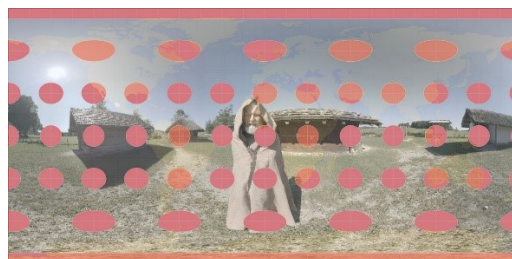


Abbildung 6-12: Äquirektanguläre Projektion mit Überlagerungen gemäß Tissotscher Indikatrix. Sie hat den Vorteil einer geringeren Verzerrung im Äquator-Bereich, aber den Nachteil einer stärkeren Verzerrung in der Nähe der Pole.

Die Heatmaps müssen unter der Annahme einer kontinuierlichen Rundumtopologie berechnet werden. Die Zellen wurden bei den Berechnungen der Heatmaps über den Rand hinweg fortgesetzt, indem die Koordinaten transformiert wurden.

Zeitabhängigkeit (Anforderung 4)

CVR-Inhalte sind von Natur aus dynamisch: Stimuli (Film), Tracks und Heatmaps ändern sich mit der Zeit. Mithilfe der Timeline können die Daten für jeden Timecode im Film entweder mit der *play*-Funktion oder der *intervall*-Funktion analysiert werden.

Bei Verwendung der *play*-Funktion ändert sich die Heatmap zeitabhängig und kann als Film angesehen werden. Ein Lifetime-Wert kann eingestellt werden, um zu spezifizieren, wie lange ein Trackpunkt in die Berechnung einfließen soll. Dieser Parameter ist für CVR von besonderer Bedeutung, da sich sinnvolle Hotspots innerhalb relativ kurzer Intervalle um einen Timecode herum bilden (Abbildung 6-13).

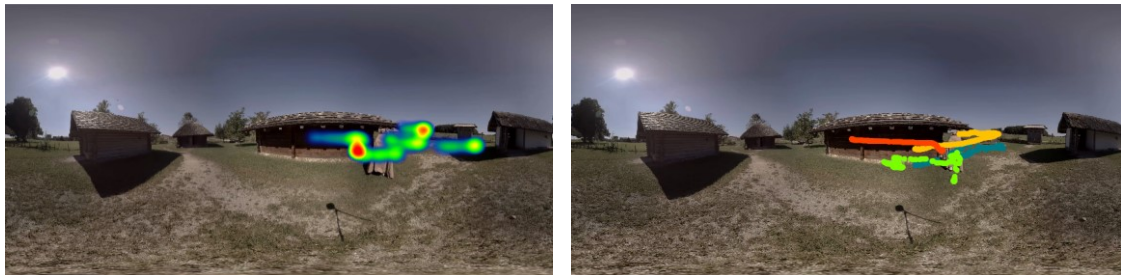


Abbildung 6-13: Darstellung von Heatmap (links) und Track (rechts) in einem Lifetime-Intervall.

Die *interval*-Funktion ist nützlich, um die Heatmap und die Tracks um einen bestimmten Timecode herum zu untersuchen, während die Wiedergabe angehalten wird.

Hotspot-Überblick (Anforderung 5)

Um die Identifizierung von Hotspots zu vereinfachen, bietet der CVR-Analyzer die Möglichkeit, das lokale Maximum der Heatvalues für jeden Timecode auf der Grundlage eines bestimmten Lifetime-Wertes vorab zu berechnen.



Abbildung 6-14: Hotspot-Erkennung auf der Timeline. Für jeden Timecode wird der höchste Heatvalue angezeigt. Auf diese Weise können Timecodes mit relevanten Hotspots erkannt werden.

Das lokale Maximum für einen Timecode ist das Maximum der berechneten Heatvalues im angegebenen Lifetime-Intervall um einen Timecode (Abbildung 6-14). Durch Klicken auf einen bestimmten Wert im Diagramm navigiert die Anwendung zum entsprechenden Timecode im Film, zeigt das Bild mit der Heatmap und markiert den Bereich des Hotspots.

Verschiedene Quellen vergleichen (Anforderung 6)

In einigen Fällen werden die Daten auf natürliche Weise gruppiert, weisen jedoch alle die gleiche Struktur auf. In unseren Experimenten haben wir verschiedene Anwendungsszenarien:

- Vergleich von Head- und Eyetracking-Daten
- Vergleich mehrerer Lenkungsmethoden unter Verwendung des gleichen Films
- Vergleich von Tracking-Daten mit sich bewegenden Stimuli (z. B. Flickern)

Für **Studie 7.1.5** benötigten wir Informationen über drei Komponenten, die sich räumlich und zeitlich verändern: Kopfdrehung, Augenbewegung und zusätzliche Stimuli, in diesem Fall Flickerregionen. Dadurch gab es drei Gruppen von Daten, die alle die gleiche räumlich-zeitliche Struktur aufwiesen.

Durch das Definieren verschiedener Gruppen von Tracking Dateien können alle oben beschriebenen Mechanismen auch auf Gruppen angewendet werden. Auf diese Weise ist es möglich, die gruppierten Daten separat zu prüfen und miteinander zu vergleichen.

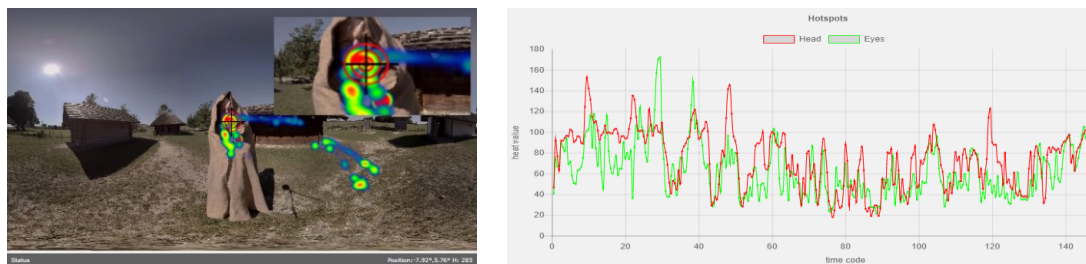


Abbildung 6-15: Markieren des Hotspots einer Gruppe in einer gemeinsamen Heatmap.

Durch Klicken auf einen bestimmten Gruppenwert im Liniendiagramm markiert die Anwendung den Bereich des Hotspots mit der entsprechenden Linienfarbe (Abbildung 6-15). Die Koordinaten werden in der Statusleiste angezeigt.

Performance

Heatmaps synchron zum Timecode des Films anzuzeigen, kann Performanceprobleme verursachen. Um eine Visualisierung auch auf Computern mit geringerer Leistung zu ermöglichen, bietet das Tool deshalb einige Parameter zur Performanceverbesserung:

- Standardmäßig wird bei jedem Aufruf des Animationsframes ein Canvas gerendert (auf unserer Hardware ca. 60 Mal pro Sekunde). Es ist möglich eine niedrigere Bildrate für Heatmap-Updates zu wählen. Dadurch wird die Anzahl der Canvas Updates pro Sekunde verringert.
- Defaultmäßig wird ein 1000x500 Pixel Canvas angezeigt. Dies führt zu einer fein strukturierten Visualisierung bei hohem Rechenaufwand. Die Auswahl einer gröberen Pixelstruktur des Canvas ist möglich. Die Verwendung eines Canvas mit 500 x 250 Pixel reduziert den Aufwand auf ein Viertel.

6.2.3 Zusammenfassung

Die Hauptfunktionen des CVR-Analyzers sind:

- Die aufgezeichneten Daten können in zwei Modi auf dem Film angezeigt werden: dem *Heatmap*- und dem *Track-Modus*.
- Mit der *play*-Funktion können sowohl Heatmap als auch Tracks auf dem laufenden Film betrachtet werden.
- Die *intervall*-Funktion ermöglicht die Analyse der Heatmap und Tracks in einem bestimmten Timecode-Intervall, während die Wiedergabe angehalten ist.

- Der *Hotspot*-Überblick zeigt die zeitliche Veränderung der Heatvalues auf einen Blick und ermöglicht es so, schnell relevante Timecodes zu finden.
- Durch die Definition von *Gruppen* können die gruppierten Daten separat überprüft und miteinander verglichen werden.

6.3 Räumliche Statistik

Mittels Punktmusteranalyse (engl.: point pattern analysis) bestimmt man, ob eine beobachtete Verteilung von Punktereignissen aus einem zufälligen Muster resultiert oder einem systematischen Prozess folgt. Im Folgenden werden zwei Methoden vorgestellt, die für Raum-Zeit-Daten geeignet sind.

6.3.1 Raumzeitwürfel

Eine Methode zur Untersuchung von Raum-Zeit-Daten ist die Analyse von Raum-Zeit-Würfeln (Space-Time Cube, STC). In diesen Würfeln repräsentieren zwei Koordinaten räumliche Werte und die dritte die Zeit. Mit dieser Technik können Raum-Zeit-Daten visualisiert und untersucht werden (Abbildung 6-16).

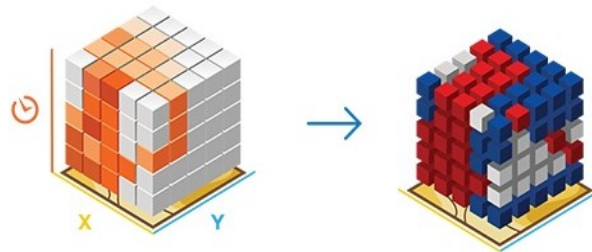


Abbildung 6-16: Ein Raum-Zeit-Würfel hat eine Zeit- und zwei Raumkoordinaten. Links wird die Dichte eines Werts für jeden Raum-Zeit-Abschnitt angezeigt, rechts das Ergebnis einer Hotspot-Analyse (signifikante Hotspots sind rot)(esri, 2018).

In unserer Arbeit haben wir für die Berechnung des STC und die Anwendung von Methoden der räumlichen Statistik die GIS-Software ArcGIS Pro genutzt. Dazu wurden die aufgezeichneten Tracking-Daten in eine Excel-Datei konvertiert und in ArcGIS importiert, um die implementierte STC-Methode für die Analyse der Daten zu verwenden.

In einem ersten Schritt wurden die Raum-Zeit-Würfel auf die Anzahl der Ereignisse untersucht – in unserem Fall, „wie oft“ Benutzer einen bestimmten Bereich innerhalb eines bestimmten Zeitintervalls betrachteten (Abbildung 6-17).

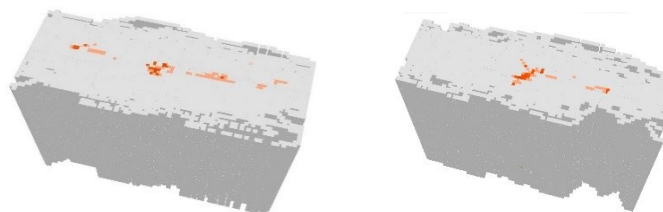


Abbildung 6-17: Beispiel für Raum-Zeit-Würfel: zwei Teilnehmergruppen, sahen den gleichen Film mit unterschiedlichen Methoden. Die orangen Regionen wurden oft angeschaut.

Auf diese Weise konnten wir für jedes Zeitintervall (in unserem Fall 1 s) identifizieren, wo die meisten Trackpunkte liegen. Die Korrelation zwischen den Nachbarpunkten (in Raum und Zeit) wurde in diesem Schritt jedoch noch nicht berücksichtigt. Dazu benötigt es eine Hotspot-Analyse.

6.3.2 Getis-Ord (Gi*) Statistik

Um statistisch signifikante Hotspots zu finden, können Methoden der räumlichen Statistik angewendet werden. Diese Methoden berücksichtigen die Nachbarschaftsbeziehungen zwischen den Raum-Zeit-Segmenten. Eine Statistik für diesen Zweck wurde von Getis und Ord entwickelt: die Getis-Ord (Gi*) Statistik (Getis und Ord, 1992; Ord und Getis, 1995). Diese statistische Methode kann eine Tendenz zur räumlichen Clusterbildung erkennen und zwischen Orten mit hohen und niedrigen räumlichen Assoziationen unterscheiden (Songchitruksa und Zeng, 2010).

Dafür wird das Untersuchungsgebiet in n Segmente unterteilt, wobei jedes Segment einem zentralen Punkt i (1, 2, ..., n) zugeordnet ist. Die Ereignisse (engl.: incidents) werden mit kartesischen Koordinaten versehen. Ziel ist es, die Existenz eines räumlichen Musters für eine Zufallsvariable X zu untersuchen. Dabei ist X ein ausgewähltes Ereignisattribut, dessen Werte x_i jedem Segment zugeordnet werden. Wenn x_i Ähnlichkeiten zwischen benachbarten Regionen aufweist, kann man behaupten, dass eine räumliche Autokorrelation der Variablen X über dem Segment i existiert.

Die in dieser Arbeit untersuchten Tracking-Daten sind Punktereignisdaten. *Punktereignisdaten* sind Koordinaten, die mit Ereignissen verbunden sind. In unserem Fall ist das Ereignis der Blick eines Betrachters auf diesen Punkt. Die Gi*-Statistik eignet sich, um *signifikante* Hotspots zu identifizieren. Dazu werden Segmente gebildet, die Ereignisdaten aggregiert und die Anzahl der Ereignisse in den Segmenten ermittelt. Die Anzahl der Ereignisse – in unserem Fall die Anzahl der Blicke in einem Segment – sind die Attributwerte, die von der Methode analysiert werden.

Die Gi*-Statistik lautet:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2}{n-1}}}$$

wobei x_j der Attributwert für Punkt j , $w_{i,j}$ das räumliche Gewicht zwischen den Punkten i und j , n gleich der Gesamtzahl der Punkte ist und:

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n}$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2}$$

Mit den räumlichen Statistik-Tools der ArcGIS Pro-Software wurden die Signifikanzlevels ermittelt (Abbildung 6-18).

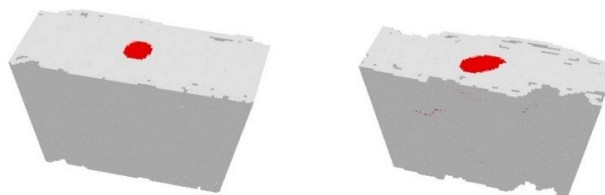


Abbildung 6-18: Die Gi*-Statistik zeigt signifikante Hotspots für beide Gruppen.

Für jeden Punkt in der Datenmenge kann der p-Wert durch Doppelklick auf den Punkt angezeigt werden. Der p-Wert repräsentiert die Wahrscheinlichkeit, dass das beobachtete Muster zufällig erzeugt wurde. Ein kleiner p-Wert bedeutet, dass das Muster höchstwahrscheinlich von einem Cluster verursacht wird. Segmente mit p-Werten kleiner als 0,01, bedeuten eine 99%ige Konfidenz und werden rot angezeigt.

Die Verwendung des 3D-Modus des GIS-Tools für eine detailliertere Untersuchung der Daten war mit unserem Datensatz sehr zeitaufwendig. Der 3D-Modus war zwar für eine Übersicht des Datensatzes geeignet, aufgrund der Größe unseres Datensatzes war es jedoch schwierig, im 3D-Würfel zu navigieren. Wir verwendeten einen Acer Predator G3-710 (Intel i7-6700, 3,4 GHz, 16 GB RAM). Das Navigieren in kleineren Datensätzen war kein Problem. Durch Wechsel in den 2D-Modus und das Ändern der Zeit mit einem Schieberegler, konnten jedoch signifikante Hotspotbereiche auf komfortable Weise gefunden werden (Abbildung 6-19).

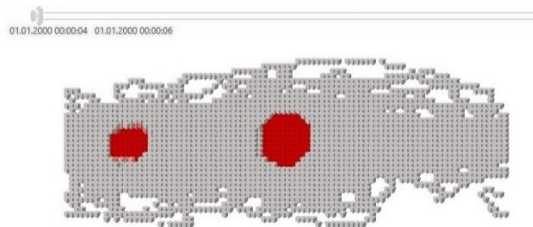


Abbildung 6-19: Mit einem Schieberegler im 2D-Modus kann man in den Daten navigieren und die Zeit ändern.

6.4 Fragebögen für Cinematic VR

Viele Fragestellungen in CVR sind eng verknüpft mit allgemeinen VR-Problemen: Es ist wichtig, ob eine untersuchte Methode die Presence beeinflusst, Simulator-Sickness hervorruft, belastend ist oder die User Experience verändert. Aus diesem Grund wurden in den Studien evaluierte Fragebögen der VR-Forschung benutzt. Evaluierte Fragebögen geben die Möglichkeit bereits gesammelte Erfahrungen zu nutzen. Einige der Fragen sind allerdings für CVR nicht relevant oder müssen angepasst werden, da manche Formen von Bewegung und Interaktion nicht existieren. Deshalb wurden diese Fragebögen in verkürzter Form angewendet. Wir sind uns bewusst, dass dadurch die Ergebnisse nicht mit anderen Studien vergleichbar sind und eine neue Evaluation der Fragebögen notwendig wäre. Für den etablierten NASA-TLX-Fragebogen wurde 20-Jahre nach dessen Einführung untersucht, wie dieser an unterschiedliche Versuchssituationen angepasst wurde (Hart, 2006). Die Ergebnisse zeigen, dass auch mit modifizierten Fragebögen wichtige Ergebnisse gefunden werden können. Modifizierten Fragebögen sind nützlich, um Methoden miteinander zu vergleichen und Informationen über deren Schwächen und Stärken herauszufinden. Für zukünftige Forschungen wären evaluierte Fragebögen für CVR erstrebenswert.

6.4.1 Presence

Zur Untersuchung der Presence wurden in unseren Studien zwei unterschiedliche Fragebögen verwendet: zum einen der Presence-Fragebogen (Presence Questionnaire – PQ) von Witmer und Singer (Witmer und Singer, 1998) zum anderen der IPQ (Schubert u. a., 2002).

Bei Verwendung des PQ, wurden folgende für CVR relevante Fragen ausgewählt:

- (PQ1) Wie sehr waren Sie in das virtuelle Erlebnis involviert? (Involvement)
- (PQ2) Wie sehr waren Sie durch die visuellen Aspekte involviert? (Visuell)
- (PQ3) Wie sehr waren Sie durch die auditiven Aspekte involviert? (Audio)
- (PQ4) Wie schnell haben Sie sich an die virtuelle Umgebung angepasst? (Zeit)
- (PQ5) Inwieweit stimmten Ihre Erfahrungen in der virtuellen Umgebung mit Ihren realen Erfahrungen überein? (Übereinstimmung)

Die Fragen werden auf einer 7-Punkte-Likert-Skala beantwortet (1= gar nicht, 7=komplett/sehr)

Bei Verwendung des IPQ, wurden je nach Studienschwerpunkt unterschiedliche Fragen ausgewählt, die in den einzelnen Studien vermerkt sind.

6.4.2 Task Workload

Die Belastung wurde unter Verwendung des NASA-TLX-Fragebogens (Hart und Staveland, 1988) untersucht, bei dem alle sechs Unterskalen verwendet wurden: (1) geistige Belastung, (2) physische Belastung, (3) zeitliche Belastung, (4) Performance, (5) Anstrengung, (6) Frustration.

Wie in (Hart, 2006) beschrieben, wurde der Gewichtungsprozess eliminiert und der Raw-TLX (RTLX) verwendet. Zusätzlich zur Gesamtlast wurden die Subskala-Scores jedes einzelnen Items verglichen, um mögliche Gründe für eine erhöhte Belastung zu finden.

6.4.3 Simulator-Sickness

Zur Messung der Simulator-Sickness wurde ein reduzierter Fragebogen des Simulator-Sickness Questionnaire (SSQ) von Kennedy et al. (Kennedy u. a., 1993) verwendet. Da nicht alle Fragen für Cinematic VR relevant sind, wurden sechs Punkte ausgewählt: (1) allgemeines Unbehagen, (2) Müdigkeit, (3) Kopfschmerzen, (4) Augenbelastung, (5) Fokussierungsschwierigkeiten, (6) Übelkeit, (7) Konzentrationsschwierigkeiten.

Für jedes Item wird vom Teilnehmer ein Sickness-Level ausgewählt (keine, leicht, mittelschwer, schwer), und die Antworten werden auf eine Skala von 0 (keine) bis 3 (schwer) transformiert (Kennedy u. a., 1993).

Durch die Auswahl war es nicht möglich, die Scores mit den Gewichtungsmethoden so zu berechnen, wie es in der Originalarbeit beschrieben ist (Kennedy u. a., 1993). Unter Verwendung der ausgewählten Elemente konnten jedoch Scores für Übelkeit, Augenmotorik und Desorientierung berechnet werden, um Methoden miteinander zu vergleichen. Zusätzlich wurden die Items im Einzelnen analysiert, um mögliche Ursachen für Simulator-Sickness zu ermitteln. Ähnlich wie bei der Verwendung des NASA-TLX wurden die Raw-Werte ohne Gewichtung verwendet.

6.4.4 User Experience und Enjoyment

Auch für User Experience und Enjoyment wurden etablierte Fragebögen genutzt, die für jede Studie separat ausgewählt wurden und deshalb in der jeweiligen Studie beschrieben werden.

6.5 Fazit und Ausblick

Mit dem CVR-Analyzer ist es möglich, Hotspots in den Tracking-Daten zu finden und dadurch das Sehverhalten in CVR besser zu verstehen. Das Tool bietet viele Optionen und Anpassungsmöglichkeiten an das jeweilige Szenario.

Um Verzerrungen durch die Projektion auf eine Ebene zu vermeiden, könnte eine Analyse der Daten in virtuellen Umgebungen von Vorteil sein. Bei HMDs ist das FoV jedoch relativ klein. Die Untersuchung der Daten mithilfe begehbare Surround-Anzeigesysteme wie CAVE (Cruz-Neira u. a., 1992) oder Allosphere (Wakefield u. a., 2013) könnte eine komfortable Ergänzung zum Untersuchen solcher Daten bieten.

Um gefundene Hotspots auf Signifikanz zu überprüfen, eignen sich Methoden der räumlichen Statistik. Mittels der Getis-Ord Gi*-Statistik ist es möglich die Signifikanz zu ermitteln. Zur Anwendung der Getis-Ord Gi*-Statistik verwendeten wir die GIS-Software ArcGIS Pro. Eine Integration dieser Methode in den CVR-Analyzer würde eine noch komfortablere Analyse ermöglichen.

Der CVR-Analyzer kann auch für Forschungen in AR und VR genutzt werden. Erste Studien in diesen Bereichen wurden bereits durchgeführt.

7

Nutzerstudien

„Aber wenn auch jeder Film, der bis jetzt gemacht wurde, schlecht und unkünstlerisch wäre, ist es nicht gerade die Aufgabe von Euch Theoretikern, seine prinzipiellen Möglichkeiten zu erforschen? Diese wären des Wissens wert, auch wenn nie eine Hoffnung für ihre Realisierung bestände. Die gute, die schöpferische Theorie ist keine Erfahrungswissenschaft und wäre vollkommen überflüssig, wenn sie warten müßte, bis die Kunst in allem vollendet schon vorhanden ist. Die Theorie ist, wenn auch nicht das Steuerruder, doch zumindest der Kompaß einer Kunstentwicklung.“

Béla Balázs (Balázs, 1924)

Die nachfolgenden Nutzerstudien wurden in den Jahren 2017- 2019 durchgeführt. Bis auf die **Studien 7.1.1** und **7.1.2** sind alle Bestandteile von Bachelor- und Masterarbeiten, welche in dieser Zeit von mir betreut wurden. In diesem Kapitel werden die Methoden und Ergebnisse der Studien vorgestellt. Informationen zur Implementierung und weitere Details zur Studiendurchführung kann man in den jeweiligen Abschlussarbeiten finden. Die Ergebnisse der Studien werden in den jeweiligen themenbezogenen Kapiteln der vorliegenden Arbeit diskutiert und eingeordnet.

Die meisten unserer Studien haben explorativen Charakter. Explorative Datenanalyse ist der Hypothesenprüfung vorzuziehen, wenn noch wenig über einen Forschungsgegenstand bekannt ist. Es gibt keine festen Annahmen und es wird unvoreingenommen beobachtet mit dem Ziel, die gewonnenen Erkenntnisse zu strukturieren (Kamps, 2018a). Dabei werden Methoden der deskriptiven Statistik verwendet, bei der die Testergebnisse durch Kennwerte (z.B. Mittelwerte), Tabellen und Grafiken beschrieben werden (Kamps, 2018b).

Im Gegensatz dazu gibt es bei Hypothesenprüfungen bereits Annahmen, die überprüft werden. Dafür sind sauber gemessene Daten und induktive statistische Verfahren (Inferenzstatistik) notwendig, um aus den Ergebnissen einer Stichprobe auf die Grundgesamtheit schließen zu können (Kamps, 2018c).

Ziel unserer Nutzerstudien war es, Erfahrungen zu sammeln, um Zusammenhänge zu erkennen und theoretische Konstrukte zu entwickeln. Deshalb wird in diesem Kapitel die deskriptive Statistik der Inferenzstatistik vorgezogen.

Die in den Studien verwendeten Fragebögen und Analysemethoden sind in Kapitel 6 ausführlich beschrieben. Es werden die folgenden Abkürzungen benutzt: MW – Mittelwert, SD – Standardabweichung (engl.: standard deviation). Einige der Studien sind in englischer Sprache veröffentlicht worden, andere wurden von vornherein mit Fragebögen in englischer Sprache durchgeführt. In diesen Fällen wurden die Fragen in den Diagrammen nicht ins Deutsche übersetzt, um Inkonsistenz zu vermeiden.

7.1 Nutzerstudien: Aufmerksamkeitslenkung

In diesem Abschnitt werden verschiedene Methoden zur Lenkung der Aufmerksamkeit vorgestellt und miteinander verglichen, die in Kapitel 2 in einen allgemeinen Kontext gebracht werden. Die **Studien 7.1.1** und **7.1.2** untersuchen visuelle und auditive Methoden der Aufmerksamkeitslenkung, die vom Inhalt des Films beeinflusst werden. Es sind filmische Methoden, die auch ohne zusätzliche Implementierungen anwendbar sind. Die Methoden in den **Studien 7.1.3** bis **7.1.6** wurden in Unity 3D implementiert und erfordern eine speziell produzierte Applikation. In **Studie 7.1.3** erhält der Zuschauer mittels Picture-in-Picture (PiP) eine Vorschau der Region of Interest (RoI) oder wird durch einen Pfeil auf das Vorhandensein einer solchen Region aufmerksam gemacht. Die **Studien 7.1.4** bis **7.1.6** untersuchen verschiedene Flicker-Methoden.

7.1.1 Studie: Diegetisches Guiding

Die Ergebnisse dieser Studie wurden vorab veröffentlicht in:

- Rothe, S., Hußmann, H., 2018. Guiding the Viewer in Cinematic Virtual Reality by Diegetic Cues, in: International Conference on Augmented Reality, Virtual Reality and Computer Graphics. Springer, Cham
- Rothe, S., Hußmann, H., Allary, M., 2017. Diegetic cues for guiding the viewer in cinematic virtual reality, in: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17. ACM, New York, USA

Die Studie ist von mir vorbereitet, implementiert und durchgeführt worden. Die Daten habe ich analysiert und mit Heinrich Hußmann und Mathias Allary diskutiert.

Um zu untersuchen, ob diegetische Cues geeignet sind, die Aufmerksamkeit des Betrachters auf sich zu ziehen, wurden folgende Fragen untersucht:

- Welche Arten von Cues lenken die Aufmerksamkeit effektiver: Ton, Licht oder Bewegungen?
- Welche Kombination von Cues Ton/Licht, Licht/Bewegung, Bewegung/Ton ist wirksam, um die Aufmerksamkeit zu lenken?

- Kann ein diegetischer Audio-Cue die Aufmerksamkeit des Betrachters auf sich ziehen, auch wenn die Quelle sich in dem Moment, in dem der Cue erscheint, nicht im FoV befindet? Wie wichtig ist die Richtung des Geräusches für die Lenkung?
- Kann ein diegetischer bewegter Cue die Aufmerksamkeit des Betrachters lenken, auch wenn er sich nicht ständig im FoV befindet?

7.1.1.1 Material

Zur Beantwortung der obigen Fragen wurde ein Film mit unterschiedlichen diegetischen Cues produziert:

- Ton aus einer bestimmten Richtung (*s-sound*),
- beleuchtete Objekte (*l-light*),
- Bewegungen von stationären Objekten - zum Beispiel Schwingen (*m-motion*),
- sich fortbewegende Objekte - Bewegungen mit Positionsänderung (*lo-locomotion*).

In den für den Test relevanten Szenen ist der Ton räumlich und mit visuellen Objekten verbunden. In den Szenen, die zur Randomisierung der Blickrichtung verwendet werden, wurde Stereoton verwendet.

Die Kopfbewegungen der Betrachter wurden aufgezeichnet, so dass beurteilt werden konnte, ob diese von den Cues beeinflusst wurden. Für den gesamten Film wurde eine dynamische Heatmap generiert (siehe Kapitel 6).

7.1.1.2 Teilnehmer und Prozedur

27 Personen (10 weiblich, 17 männlich) nahmen an der Studie teil und sahen einen CVR-Film mit einem HMD (Samsung Galaxy S6 mit Google Cardboard) und Kopfhörern (Ultrasone HFI 780). Die Teilnehmer waren zwischen 16 und 73 Jahre alt (MW=40). Ein Teilnehmer war Tontechniker und untersuchte während der Studie, wie der räumliche Schall in der Testumgebung auf ungewöhnliche Kopfbewegungen reagiert. Im abschließenden Interview stellten wir fest, dass seine Aufmerksamkeit nicht auf den Film gerichtet war und entfernten seinen Datensatz.

Es wurde ein within-subject Testdesign angewendet - alle Teilnehmer sahen sich den gleichen Film mit den gleichen Cues an. Es gab keine spezielle Aufgabe, stehend konnten sich die Teilnehmer umherdrehen, um die Objekte zu betrachten, an denen sie interessiert waren.

Nach dem Film wurde ein kurzes Interview geführt. Die Teilnehmer zählten die Objekte auf, an die sie sich spontan erinnern konnten. Anschließend wurde für die verbliebenen Objekte gefragt, ob diese gesehen wurden. Diese Studie untersuchte nicht primär den Einfluss auf die kognitive Wahrnehmung, jedoch halfen diese zusätzlichen Informationen bei der Interpretation der aufgezeichneten Headtracking-Daten.

7.1.1.3 Film

Der Film besteht aus 4 Abschnitten. Die Abschnitte sind durch eine neutrale Szene unterteilt: ein Wald, der in jeder Richtung ähnlich aussieht (Abbildung 7-1). Diese Szene ermöglicht eine Streuung der Blickrichtungen, so dass diese am Anfang jeder Szene zufällig verteilt sind.



Abbildung 7-1: Der Wald trennt die Testszenen, um die Blickrichtung für den Beginn der nächsten Szene zu randomisieren.

Alle anderen Szenen wurden am selben Ort aufgenommen - einer mysteriösen Küche eines alten Schlosses (Abbildung 7-2). Objekte mit unterschiedlichen Cues erscheinen und verschwinden von Zeit zu Zeit. Der Ton in diesen Szenen ist räumlich und mit den erscheinenden Objekten verbunden. Zum Verlinken der Sounds mit den Objekten wurde Unity 3D verwendet.



Abbildung 7-2: Der Raum der Szene.

Im ersten Teil des Films (Szene 1 und 2) ändern die Objekte ihre Position nicht. Objekte mit Bewegungen schwingen oder flickern. Andere Objekte werden beleuchtet oder mit Tönen verbunden. Alle Objekte befinden sich vom Anfang bis zum Ende in der Szene. Es wurde untersucht, ob diese Objekte die Aufmerksamkeit des Betrachters auf sich ziehen können.

Ziel des zweiten Teils (Szene 3 und 4) war es, zu untersuchen, ob die Blickrichtung des Teilnehmers durch Objekte verändert werden kann, die sich anfangs nicht unbedingt im Sichtfeld befinden. In Szene 3 werden dazu Objekte mit Geräuschen genutzt, in Szene 4 sich bewegende Objekte.

In der ersten Szene sind die untersuchten Objekte nur mit einem Cue verbunden. Es gibt

- ein sich bewegendes Objekt (*m*),
- ein Objekt mit räumlichen Geräuschen (*s*) und
- ein beleuchtetes Objekt (*l*) im Raum.

In der zweiten Szene wird jedes Objekt mit zwei Cues versehen:

- ein beleuchtetes bewegliches Objekt (*l, m*),
- ein sich bewegendes Objekt mit räumlichem Klang (*m. s*),
- ein beleuchtetes Objekt mit räumlichem Klang (*l, s*).

Somit sind alle Kombinationen der drei Methoden in der Szene vorhanden. Abbildung 7-7 zeigt die Anordnung der Objekte.

In der dritten Szene wurde untersucht, ob es möglich ist, die Blickrichtung mithilfe von Objekten mit Geräuschen zu ändern. Dazu erscheinen und verschwinden mehrere Objekte mit und ohne Ton in der Küche. Zu Beginn gibt es einen tickenden Wecker, gefolgt von einem pfeifenden Teekessel und einer Kuckucksuhr – alle an verschiedenen Positionen. Anschließend erscheinen zwei Telefone - nur eines davon mit Ton. Nachdem diese wieder verschwunden sind, sind in der Küche ein Grammophon (mit Ton) und ein alter Plattenspieler (ohne Ton) zu sehen - gefolgt von einem alten Radio. Die Objekte befinden sich an mehreren Positionen (Abbildung 7-3) und nicht immer im FoV der Teilnehmer. Alle Geräusche sind räumlich.

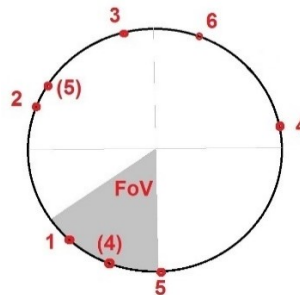


Abbildung 7-3: Die Anordnung der Objekte im Raum, die Zahlen geben die Reihenfolge des Erscheinens an: 1-Wecker, 2-Kessel, 3-Kuckucksuhr, 4 – Telefon, (4) – geräuschloses Telefon, 5-Grammophon, (5) - geräuschloser Plattenspieler.

Die vierte Szene untersucht sich bewegende Objekte, die ihre Position (*lo*) mit oder ohne zusätzliche Cues ändern. Es wurden drei verschiedene Methoden getestet, um den Blick in eine andere Richtung zu lenken:

- bewegtes Objekt ohne Ton - Feder (*lo*),
- bewegtes Licht ohne Ton - Lichtkegel (*lo*, *l*),
- bewegtes Objekt mit Raumklang - Biene (*lo*, *s*).

7.1.1.4 Ergebnisse

In einem ersten Schritt wurden die Heatmaps auf relevante Timecodes im Film untersucht und Hot Spots, auf ihre Signifikanz überprüft. Dafür wurden Methoden der räumlichen Statistik verwendet.

Zur Anwendung der Getis-Ord (Gi*) Statistik wurde die GIS-Software ArcGIS Pro genutzt und für jeden Punkt im Ereignisdatensatz der p-Wert berechnet. Der p-Wert ist die Wahrscheinlichkeit, dass das beobachtete Muster zufällig erzeugt wurde. Ein kleiner p-Wert bedeutet, dass das Muster höchstwahrscheinlich von einem Hotspot verursacht und nicht zufällig ist. Segmente mit p-Werten kleiner als 0.01, bedeuten eine 99%ige Konfidenz und werden rot angezeigt. Abbildung 7-4 zeigt die Farbe für die verschiedenen Konfidenzintervalle. Diese Legende wird zur Interpretation der Ergebnisse im nächsten Abschnitt verwendet.

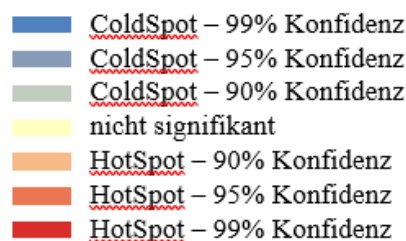


Abbildung 7-4: Legende für den Signifikanztest der Hotspots.

Szene 1

In der ersten Szene wurden drei Objekte mit einem Cue untersucht: Steckdose (*l*), Kelle (*m*), tropfendes Wasser (*s*). Abbildung 7-5 zeigt die Positionen der Objekte in dieser Szene.



Abbildung 7-5: Objekte mit einem Merkmal: Steckdose - Licht (l), Kelle - Bewegung (m), Wasser-Sound (s).

Wir konnten für keines der markierten Objekte eine größere Aufmerksamkeit beobachten. Wie in Abbildung 7-6 zu sehen ist, waren die Hotspots mit Ausnahme des Bodens und der Decke über den gesamten Raum verteilt.

Der Signifikanztest (Abbildung 7-6, rechts) zeigte mehrere Hotspots auf. Um das Soundobjekt und das sich bewegende Objekt entstanden jeweils ein signifikanter Hotspot. An der Position des Lichts (das sich nicht bewegte) gab es keinen signifikanten Hotspots. Stattdessen bildeten sich Hotspots auf anderen Objekten.

Die Szene dauert 17 s (Timecode 0:20-0:37). Im Interview wurden Steckdose und Wasser von nur 3 Personen aufgelistet. Dagegen erwähnten 16 Teilnehmern die Kelle, die allerdings bis zum Ende des Films an dieser Stelle hin- und herschwang. Bei einer detaillierten Überprüfung der Tracking-Daten haben wir festgestellt, dass die meisten Menschen die Kelle später für eine längere Zeit inspizierten, allerdings nicht in der ersten Szene.

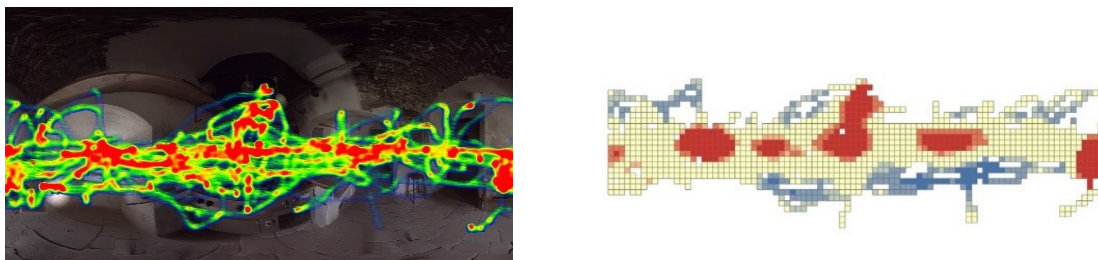


Abbildung 7-6: In Szene 1 wurden viele kleine Hotspots erzeugt.

Szene 2

Die zweite Szene hat drei Objekte mit jeweils zwei Cues: eine flickernde Lampe (m,l), ein Topf auf dem Feuer (s,m), und eine tickende Uhr (l,s). Die Anordnung der Objekte ist in Abbildung 7-7 zu erkennen.

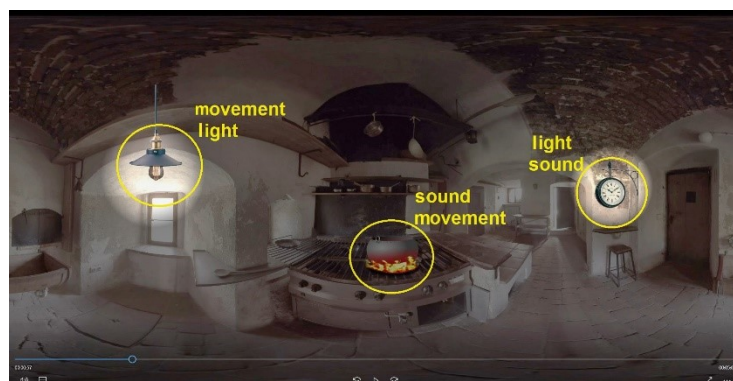


Abbildung 7-7: Objekte mit zwei Cues: Lampe (m,l), Topf (s,m), Uhr (l,s).

Wie in Abbildung 7-8 zu sehen ist, entstanden um die Objekte mit Geräuschen (Feuer, Uhr) zwei signifikante Hotspots. Die leuchtende flackernde Lampe (ohne Ton) wurde von den Teilnehmern nicht wahrgenommen. Neben dem fehlenden Ton kann dies an der Position der Lampe gelegen haben. Sie befindet sich höher als die anderen Objekte.

Die Szene dauerte 18 s (0:49-1:07). Kein Teilnehmer erwähnte die Lampe, das Feuer wurde von 18 Personen und die Uhr von 12 Personen aufgelistet.

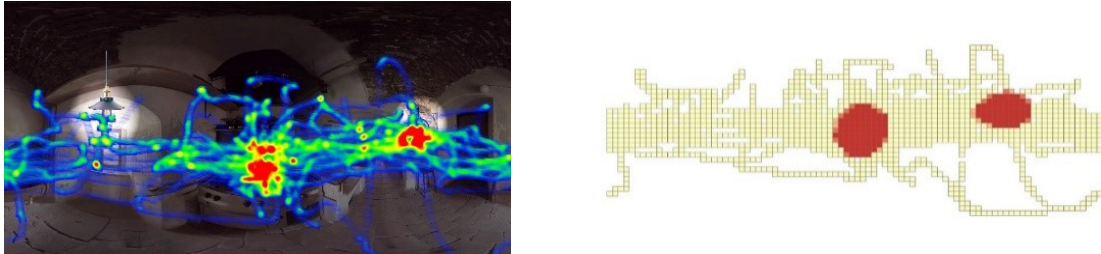


Abbildung 7-8: Es gibt Hotspots am Feuer und an der Uhr, aber keinen Hotspot an der Lampe.

Bei der Untersuchung der Tracking-Daten zu Beginn der Szene (0:50-0:52) stellten wir fest, dass die meisten Teilnehmer zuerst auf das Feuer schauten.

Szene 3

In der dritten Szene erscheinen und verschwinden mehrere Objekte mit räumlichem Sound. Sie beginnt mit einem tickenden Wecker, der 10 s später durch einen pfeifenden Teekessel und nach weiteren 10 s durch eine Kuckucksuhr an unterschiedlichen Positionen ergänzt wird (Abbildung 7-9).



Abbildung 7-9: Position der ersten drei Soundobjekte in Szene 3.

Für jedes dieser Objekte haben die meisten Teilnehmer die Blickrichtung geändert, als der Ton hörbar wurde. Wie in Abbildung 7-10 und Abbildung 7-11 zu sehen ist, wurden in allen drei Fällen signifikante Hotspots erzeugt.

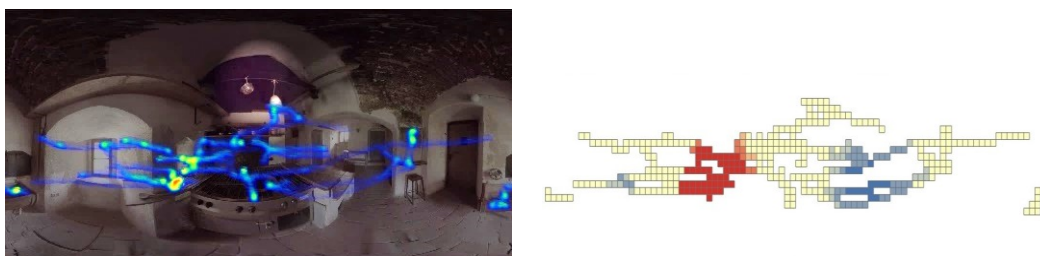


Abbildung 7-10: Heatmap und Hotspot für den Wecker.

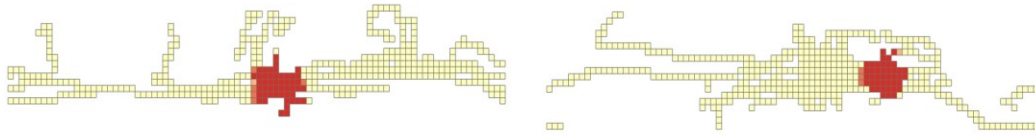


Abbildung 7-11: links: Hotspot für den Pfeifkessel, rechts: Hotspot für die Kuckucksuhr.

Danach erscheinen zwei Telefone - nur eines mit Ton (Abbildung 7-12). Um beide Objekte hat sich ein Hotspot gebildet, obwohl nur ein Telefon klingelte (Abbildung 7-13).



Abbildung 7-12: Zwei Objekte: Telefon mit Sound (rechts), Telefon ohne Sound (links).



Abbildung 7-13: Heatmap und Hotspots für die Telefone mit und ohne Sound.

Bei der Analyse der Tracking-Daten haben wir festgestellt, dass das Telefon mit Ton eine stärkere Wirkung hat. Im Interview haben mehr Personen das klingelnde Telefon in der Nähe der Tür aufgelistet (11 Teilnehmer) als das Telefon ohne Ton in der Nähe des Fensters (8 Teilnehmer). Sieben Teilnehmer bemerkten beide Telefone.

Im nächsten Schritt wurde dieses Experiment mit anderen Objekten wiederholt: einem Grammophon (das eine alte Melodie spielt) und einem alten Plattenspieler (ohne Ton). Selbst wenn beide Objekte signifikante Hotspots verursachten (Abbildung 7-14), schauten diesmal mehr Teilnehmer auf das Soundobjekt als im Test zuvor. Ein Teilnehmer erwähnte, dass es sich für ihn um einen Lernprozess handelte. Er hatte beide Telefone im Test davor erkannt und suchte deshalb diesmal das Objekt mit Ton.

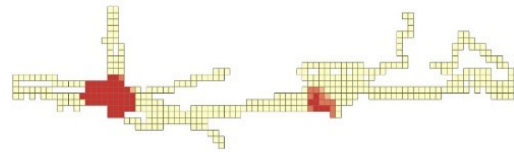
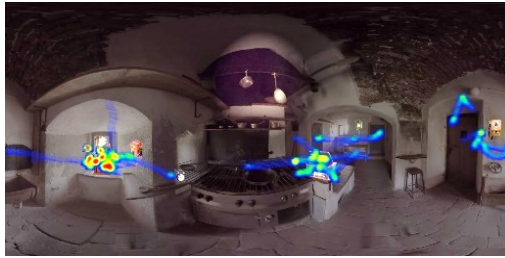


Abbildung 7-14: Grammophon mit Ton (links), Plattenspieler ohne Ton (rechts), 2 s nach dem Erscheinen der Objekte.

Im letzten Schritt dieses Testabschnittes haben wir den Winkel zum folgenden Objekt auf fast 180° erhöht (Abbildung 7-3). Gegenüber dem Grammophon erschien ein Radio. Diesmal dauerte es länger, bis sich der Hotspot bildete (Abbildung 7-15 und Abbildung 7-16).

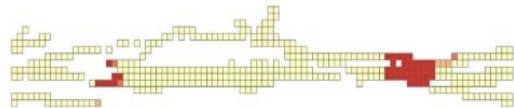


Abbildung 7-15: Heatmap und Hotspots für das Radio – 2 s nachdem das Radio erschien.

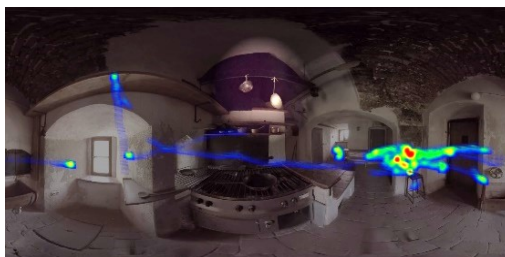


Abbildung 7-16: Heatmap und Hotspots für das Radio – 5 s nachdem das Radio erschien.

Szene 4

Die vierte Szene enthält drei sich bewegende Objekte (*lo*), zwei davon mit zusätzlichen Cues (*l*, *s*):

- eine Feder (*lo*)
- ein sich bewegender Lichtkegel (*lo*, *l*)
- eine summende Biene (*lo*, *s*).

Wenn ein Objekt eine Zielposition erreicht, erscheint ein Musikinstrument und beginnt zu spielen. Das sich bewegende Objekt verschwindet. Die Musikinstrumente bleiben vorhanden und spielen bis zum Ende der Szene. Auf diese Weise wird die Szene immer komplexer - visuell und akustisch.



Abbildung 7-17: Sich bewegende Feder. Die Feder schwebt vom Startpunkt (SP) zum Endpunkt (EP) - es dauert 20 s.

Die Szene beginnt mit einer sich bewegenden Feder (Abbildung 7-17). Die meisten Teilnehmer folgten diesem Cue, obwohl das Zeitintervall relativ lang war (20 s). In Abbildung 7-17 ist der Abstand vom Startpunkt (SP) zum Endpunkt (EP) dargestellt. Die Heatmap und die Signifikanz der Hotspots sind in Abbildung 7-18 zu sehen.

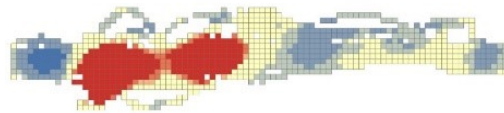
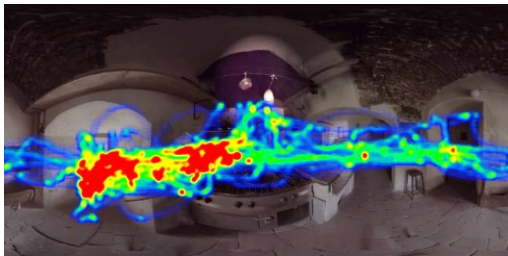


Abbildung 7-18: Heatmap und Hotspots für die sich bewegende Feder.

Für den nächsten Test wurde eine größere Entfernung und ein längeres Zeitintervall gewählt (47 s). Ein Lichtkegel bewegt sich zur gegenüberliegenden Seite des Raumes (Abbildung 7-19).

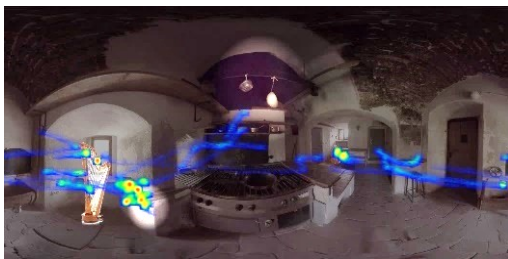


Abbildung 7-19: Der sich bewegende Lichtkegel - ein Hotspot zu Beginn des Testintervalls.

Zu Beginn des Testintervalls gab es nur einen Hotspot (Abbildung 7-19). Später (nach ungefähr 10 s) sahen sich die Teilnehmer um, vor allem zu den musizierenden Instrumenten. Es wurden drei Hotspots gebildet (Abbildung 7-20). Trotzdem kehrten die Teilnehmer meist zum Lichtkegel zurück und am Ende der Bewegung gab es wieder nur einen Hotspot (Abbildung 7-21).

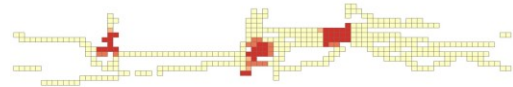


Abbildung 7-20: Der sich bewegende Lichtkegel - drei einzelne Hotspots in der Mitte des Testintervalls.

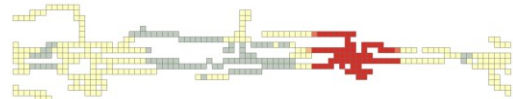
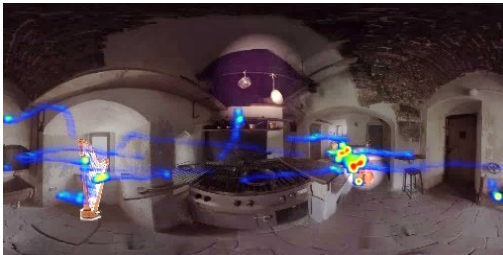


Abbildung 7-21: Der sich bewegende Lichtkegel – ein Hotspot am Ende des Testintervalls.

Der letzte Test - eine sich bewegende, summende Biene - war kürzer (10 s) und auch die zurückgelegte Strecke geringer. Infolgedessen war es schwieriger zu erkennen, ob die Teilnehmer der Biene wirklich folgten oder auf die Instrumente schauten, da die Biene von einem Instrument zum anderen flog. Selbst wenn ein Hotspot gebildet wurde, gibt es in diesem Bereich nicht nur die Biene, sondern auch die Instrumente (Abbildung 7-22).

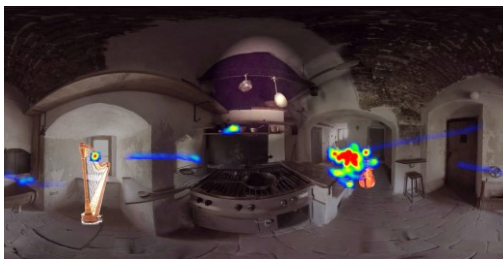


Abbildung 7-22: Ein Hotspot um die sich bewegende summende Biene, oder. die Instrumente.

Folglich wurden für alle drei in Szene 4 untersuchten Methoden signifikante Hotspots um die Cues gebildet.

7.1.1.5 Diskussion

Die vorgestellten Ergebnisse zeigen, dass diegetische Cues hilfreich sind, um die Aufmerksamkeit der Betrachter in CVR zu lenken.

In unserem ersten Experiment (Szene 1), bei dem die einzelnen Cues gleichzeitig auftraten, konnten wir keinen Cue-Typ finden, der effizienter war als die anderen. Die Hotspots waren im gesamten Raum verteilt (Abbildung 7-6). Ein Grund dafür könnte sein, dass die Cues zu subtil und nicht so auffällig waren, wie in den darauffolgenden Szenen. Außerdem sind zu Beginn des Films alle Objekte neu und Cues zeigen deshalb weniger Wirkung.

Beim Vergleich der Kombinationen von zwei Cues im zweiten Experiment (Szene 2) wurden für die Kombinationen mit Sound (s, m) und (l, s) Hotspots gefunden, allerdings wurde die Lampe (m, l) von den Zuschauern nicht bemerkt. Es bedarf weiterer Tests, ob dieses Ergebnis durch die Position des Objekts beeinflusst wurde, die höher als die der anderen war. In unserer Studie hatte die Kombination aus Bewegung und Sound die größte Wirkung, um die Aufmerksamkeit zu lenken.

Bei der Analyse der Daten aus Szene 3 konnten wir feststellen, dass Objekte, die mit Geräuschen verbunden sind, die Aufmerksamkeit des Betrachters auf sich ziehen können, auch wenn sie sich im Moment des Geräuschbeginns nicht im Sichtfeld befinden. Diese Methode ist effektiver, wenn der Sound aus der Richtung des Objekts kommt. Es funktioniert jedoch auch in anderen Fällen. Daher scheint der Klang selbst wichtiger zu sein als die Richtung des Klangs.

Darüber hinaus können bewegte Objekte die Blickrichtung beeinflussen (Szene 4). Wenn die Bewegung über einen längeren Zeitraum andauert, schauen sich die Zuschauer in der Zwischenzeit um, kehren aber meist zur Bewegung zurück. In unserem Experiment lenkten die bewegten Objekte und Lichter die Aufmerksamkeit auch ohne zusätzlichen Ton.

Zusammenfassend wurden folgende Ergebnisse gefunden:

- Objekte, die mit Sound verbunden sind, ziehen mehr Aufmerksamkeit auf sich als Objekte ohne Sound.
- Mit Sound verbundene Objekte können die Blickrichtung lenken, auch wenn der Ton nicht räumlich ist oder sogar aus einer anderen Richtung kommt.
- Bewegte Objekte oder Lichter können die Blickrichtung auch ohne Geräusche lenken.
- Es ist schwierig, den Betrachter am Anfang einer neuen Szene zu lenken.
- Nicht bewegte Lichter hatten in unseren Tests keine Wirkungen.

Um diese Ergebnisse zu finden, verwendeten wir Heatmaps und Methoden der räumlichen Statistik zur Bestimmung signifikanter Hotspots. Im anschließenden Interview haben wir erfahren, dass manche Menschen die Befürchtung haben, wichtige Ereignisse zu verpassen. Daher kann das Lenken hilfreich sein, um die CVR-Applikation angenehmer zu gestalten.

7.1.1.6 Einschränkungen

In unseren Tests wurden die Kopfbewegungen aufgezeichnet - nicht die der Augen. Dies war ausreichend für einen ersten Ansatz, und die Experimente führten zu verwertbaren Daten. Die Nutzung von Eyetracking würde allerdings noch genauere Daten liefern.

In der ersten Szene unserer Tests erregten die Cues nicht mehr Aufmerksamkeit als andere Objekte im Raum. Wir sind uns nicht sicher, ob die Ursache dafür eine Schwäche der Cues oder die Neuheit des Raumes ist. In weiteren Studien sollte die erste Szene am Ende der Testreihe wiederholt werden.

Im letzten Teil unseres Experiments, in dem eine Biene eine kurze Strecke zurücklegt, können wir nur einen Hotspot in der Mitte des Fluges sehen. Die Biene flog zwischen zwei Musikinstrumenten, die zu dieser Zeit sichtbar waren, und so fixierten die meisten Teilnehmer die Richtung des Kopfes zwischen den Instrumenten. Der Flug der Biene war kurz und die Zuschauer konnten ihm folgen, indem sie ihre Augen und nicht ihre Köpfe bewegten - so blieben sowohl die Biene als auch die Instrumente im FoV. In diesem Fall wäre Eyetracking für die Analyse hilfreicher.

Wir haben uns bemüht, den Objekten äquivalente Eigenschaften zu geben. Manchmal gab es jedoch Auswirkungen, die wir nicht erwartet hatten. Es scheint, dass Objekte, die etwas höher liegen, nicht mit Objekten gleichgesetzt werden können, die geradeaus oder etwas tiefer liegen. Dies könnte daran liegen, dass wir im wirklichen Leben seltener nach oben schauen.

Auch wenn wir die Blickrichtung zwischen den Abschnitten randomisiert haben, hingen die Experimente innerhalb einer Szene voneinander ab. In unserem Experiment haben wir die Sequenz der Cues getestet. Bei der Analyse der Daten fanden wir plausible zusätzliche Annahmen, die in weiteren Experimenten, in denen die Cues unabhängig sind, überprüft werden sollten:

- Objekte mit räumlichem Klang seitwärts des Betrachters ziehen die Aufmerksamkeit schneller auf sich als Objekte, die hinter dem Betrachter liegen
- Höhere Objekte ziehen weniger Aufmerksamkeit auf sich als Objekte auf Augenhöhe oder darunter.

7.1.1.7 Fazit

Diese Studie war ein erster Schritt in der Untersuchung von Aufmerksamkeitslenkung in CVR. Es wurde gezeigt, dass der Ton die Aufmerksamkeit des Betrachters auf sich zieht, aber auch Objekte ohne Ton zur Lenkung genutzt werden können. Die Teilnehmer folgten einem sich bewegenden Lichtkegel und anderen bewegten Objekten.

Diese Ergebnisse können verwendet werden, um diegetische Cues in einen Film zu integrieren, und damit den Betrachter zu Dingen zu führen, die für die Geschichte wichtig sind. Allerdings ist die Integration von Cues in einen Film nicht immer möglich. Deshalb wurden weitere Studien durchgeführt, um auch nicht-diegetische Methoden für die Zuschauerlenkung zu untersuchen (7.1.3 - 7.1.6).

7.1.2 Studie: Vergleich spatial Sound – nonspatial Sound

Die Ergebnisse dieser Studie wurden vorab veröffentlicht in:

- Rothe, S., Rieger, M., 2018. Spatial vs. Static Sound - the Differences it Makes for 360° Videos, Highlighted by Heatmaps, in: 30th Tonmeistertagung 2018

Die Studie wurde von mir vorbereitet und implementiert. Ein Teil der Studie (spatial Sound) wurde von Martin Rieger durchgeführt, der andere Teil (nonspatial Sound) von mir. Die Daten habe ich analysiert und mit Martin Rieger diskutiert.

In dieser Studie untersuchen wir das Sehverhalten beim Schauen von Filmen mit *spatial* (räumlich) und *nonspatial* (nicht räumlich) Sound. Im Gegensatz zu herkömmlichen Filmen ist es unklar, ob die Tonquelle im FoV des Betrachters liegt, da dieses von der Betrachtungsrichtung abhängt. Es soll herausgefunden werden, in welcher Weise der Ton die Aufmerksamkeit des Betrachters auf sich zieht und ob es einen Unterschied gibt, einen Film mit räumlichem oder nicht räumlichem Klang anzusehen.

7.1.2.1 Material

Für diese Studie verwendeten wir einen 9:26-minütigen CVR-Film (Vtopia360, 2017), der in zwei Varianten produziert wurde: mit spatial und mit nonspatial Sound. Er enthält diegetische Cues, um den Betrachter zu lenken, zum Beispiel sprechende Personen, Bewegungen und Geräusche.

Die Teilnehmer sahen sich den Film über ein HMD (Samsung Gear VR mit Samsung Galaxy S6) mit Kopfhörern an und saßen dabei auf einem Drehstuhl. Nach dem Anschauen des Films folgte ein kurzes unstrukturiertes Interview.

7.1.2.2 Teilnehmer und Prozedur

Die Studie basierte auf einem between-subject Testdesign: Der Film wurde zwei Gruppen gezeigt. Die erste Gruppe (20 Teilnehmer, 10 weiblich, 10 männlich, Alter: MW = 31.5) sah den Film mit räumlichem Sound. In dieser Studie nennen wir diese Gruppe *spatial*-Gruppe. Die andere Gruppe – die *nonspatial*-Gruppe (20 Teilnehmer, 8 weiblich, 12 männlich, Alter: MW = 34.7) hörte normalen Stereoton, der nicht mit der HMD-Bewegung verbunden war. Jede Gruppe bestand aus 6 Teilnehmern, die sich zum ersten Mal CVR-Filme anschauten, und 3 erfahrenen Teilnehmern. Die übrigen Teilnehmer sahen sich gelegentlich CVR-Filme an. Es gab keine spezielle Aufgabe für die Teilnehmer.

7.1.2.3 Ergebnisse

Bei der Analyse der getrackten Kopfbewegungen fällt auf, dass die Bewegungen in beiden Gruppen sehr eingeschränkt sind. Wir haben die Gier- und Nickwinkel der Kopfbewegungen separat untersucht. Die Gierwinkel (Drehen des Kopfes zur Seite) waren für beide Gruppen ähnlich. Meistens schauten die Teilnehmer geradeaus und drehten sich selten um (Abbildung 7-23). Diese Ergebnisse ähneln denen von Pausch et al. (Pausch u. a., 1996).

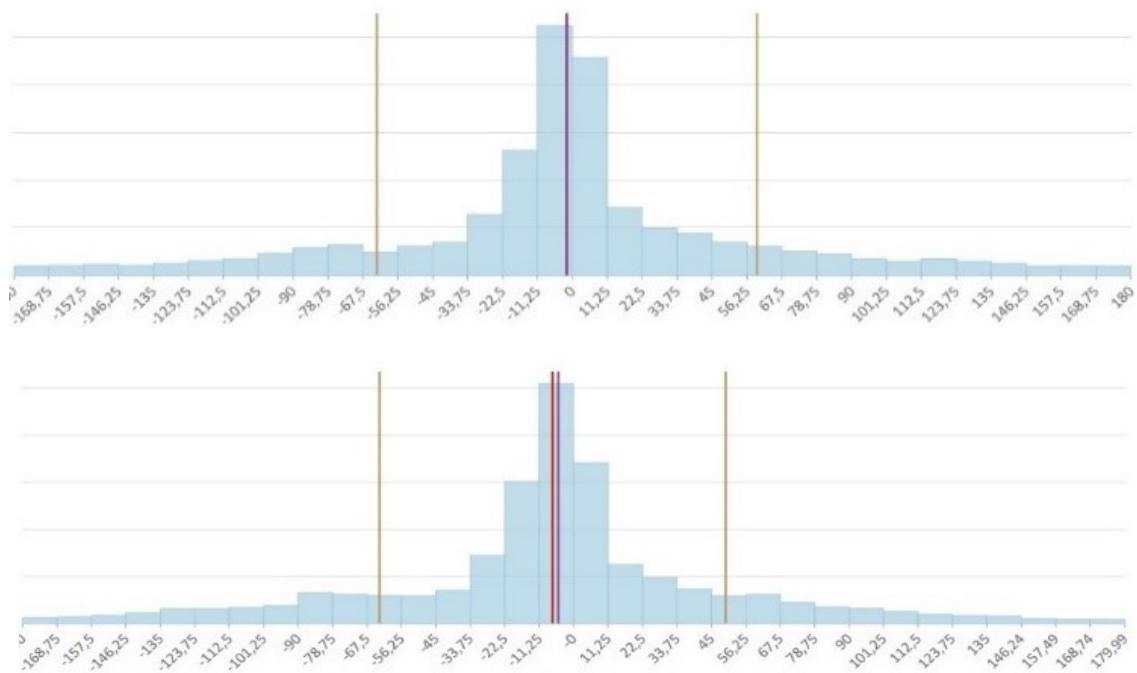


Abbildung 7-23: Die Teilnehmer schauten in beiden Gruppen kaum nach links und rechts, oben: räumlicher Sound, unten: nicht-räumlicher Sound (Linien: rot - Mittelwert, blau - Median, gelb - Standardabweichung), die x-Achse zeigt den horizontalen Winkel der Kopfbewegungen und die y-Achse die Häufigkeit (die Kopfbewegungen wurden 60 mal pro Sekunde aufgezeichnet).

Außerdem bewegten die Teilnehmer ihren Kopf kaum nach unten oder oben (Abbildung 7-24). Meist hatten die Teilnehmer den Kopf nach vorne gerichtet, kaum mehr als 15° nach unten oder 5° nach oben.

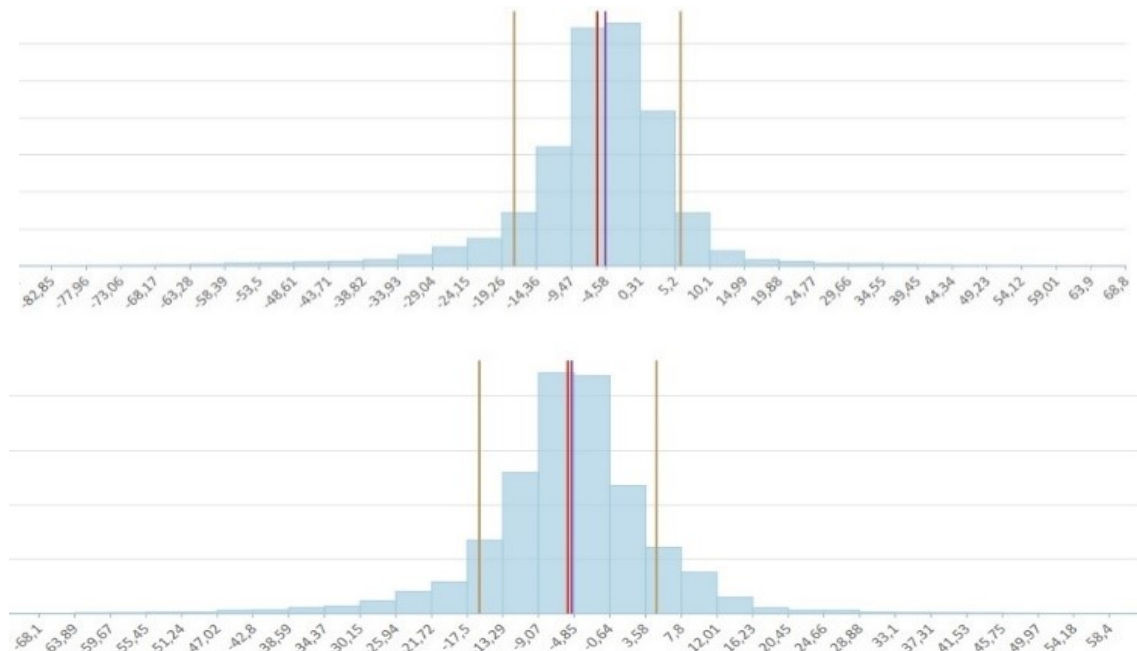


Abbildung 7-24: Die Teilnehmer sahen in beiden Gruppen kaum auf und ab, oben: räumlicher Sound, unten: nicht-räumlicher Sound (Linien: rot - Mittelwert, blau - Median, gelb - Standardabweichung), die x-Achse zeigt den vertikalen Winkel von Kopfbewegungen und die y - Achse die Häufigkeit (die Kopfbewegungen wurden 60 mal pro Sekunde aufgezeichnet).

Ein Grund dafür ist die Art unserer Daten – es wurden die Kopfbewegungen getrackt, nicht die Augenbewegungen. In der Nähe der horizontalen Linie kann der Kopf leicht bewegt werden, um die Blickrichtung zu ändern. Für den Blick auf den Boden oder den Himmel ist es jedoch bequemer, den Kopf nur geringfügig zu bewegen und die Blickrichtung durch Augenbewegungen zu erweitern. Der Bereich, in dem eine Person den Kopf mühelos bewegt, liegt bei 45° in beide Richtungen (Dreyfuss, 1967). Ein weiterer Grund für den eingeschränkten Bereich ist der Aufbau des Films - fast alle Rols befinden sich in der Nähe der horizontalen Linie.

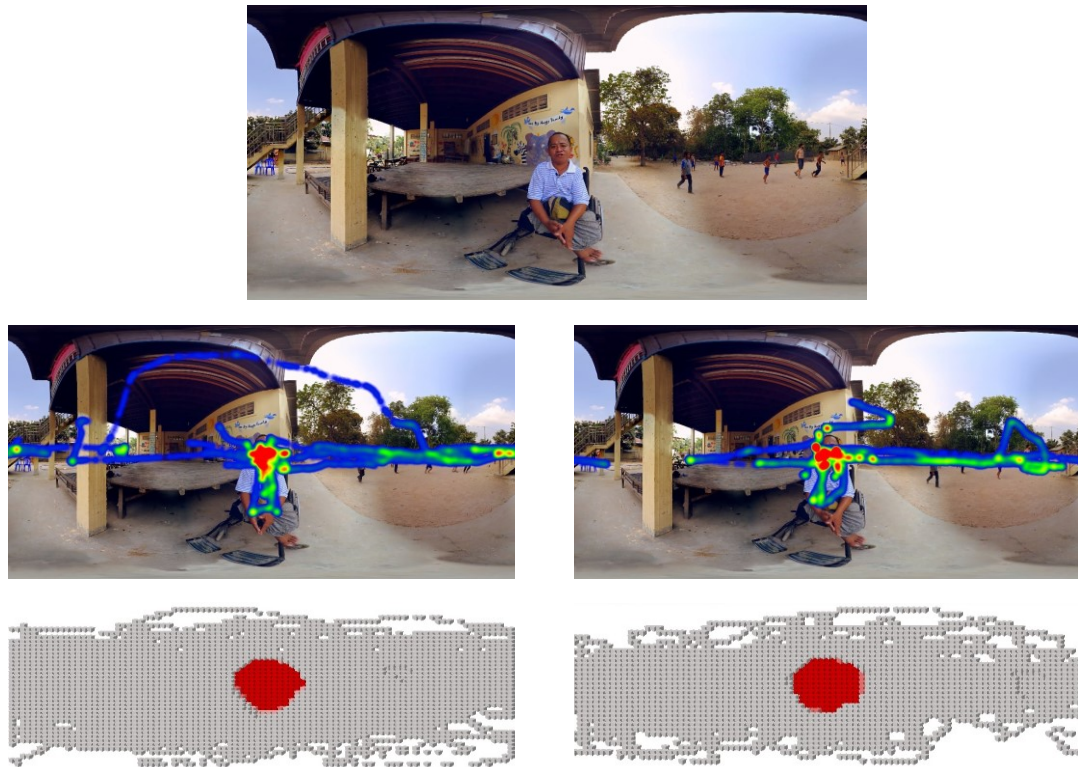


Abbildung 7-25: Sprechen oder Bewegen von Personen erregten die Aufmerksamkeit beider Gruppen in ähnlicher Weise (links: räumlicher Sound, rechts: nicht räumlicher Sound), oben: Filmszene, Mitte: Heatmaps, unten: signifikante Hotspots.

Auf den ersten Blick war das Verhalten in beiden Gruppen sehr ähnlich. Wenn Personen sprachen oder sich bewegten, gab es für beide Gruppen Bereiche mit hoher Datendichte in diesen Regionen (Abbildung 7-25).

Unter Verwendung der Getis-Ord Gi*-Statistik zur Identifizierung von Hotspots fanden wir in den Daten der Gruppe, die den Film mit nicht-räumlichem Sound angesehen hat, mehr Hotspots (99 Hotspots in der *spatial*-Gruppe und 149 Hotspots in der *nonspatial*-Gruppe). Anschließend wurde zur genaueren Betrachtung durch die Zeitkoordinate navigiert. Auf diese Weise konnte ein weiterer Unterschied zwischen den beiden Gruppen festgestellt werden. Wenn in einer Szene zwei oder mehr Cues gleichzeitig auftraten (zum Beispiel: Sprechen an einem Ort und Bewegungen an einem anderen), entstanden in der *nonspatial*-Gruppe zusätzliche Hotspots. Abbildung 7-26 zeigt zum Beispiel, dass sich zusätzlich zu den Hotspots um die Mopeds ein Hotspot um das Schild bildete.

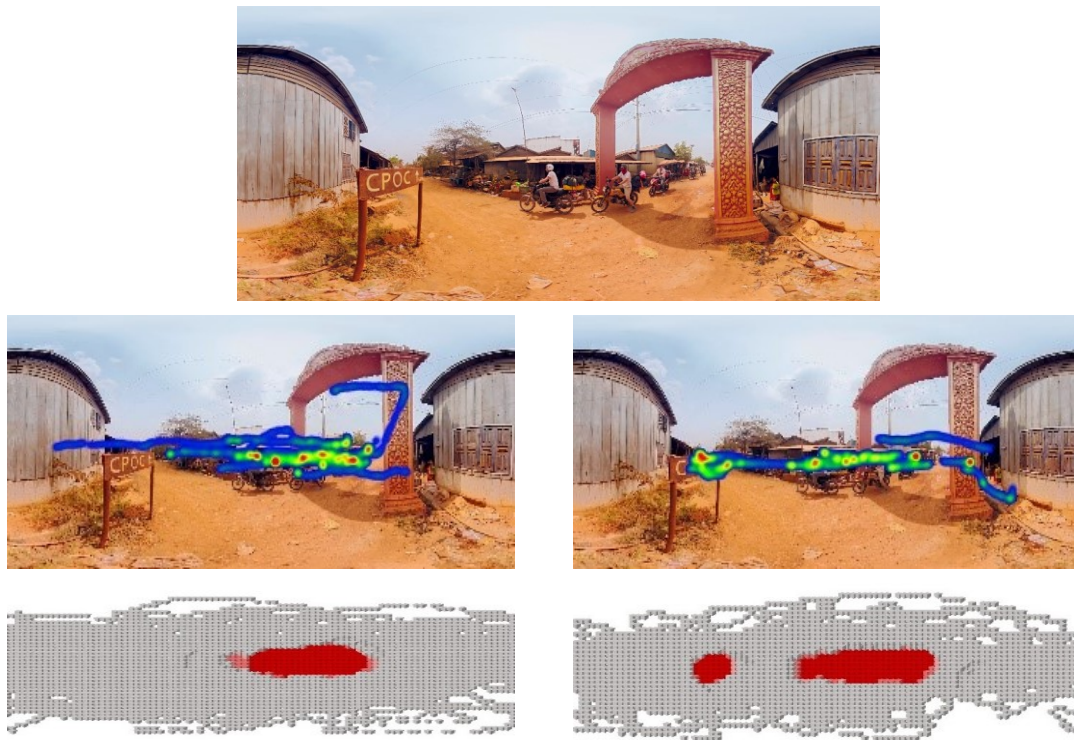
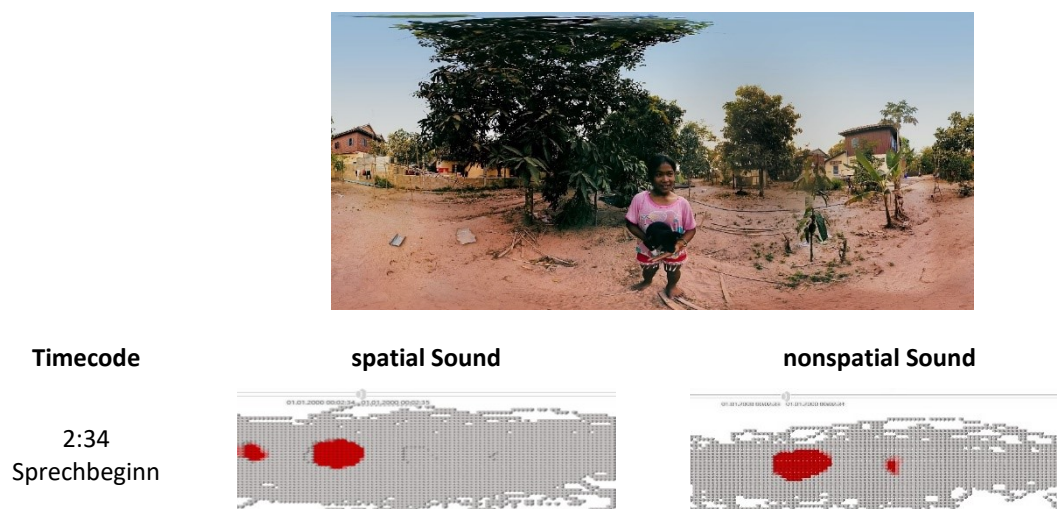
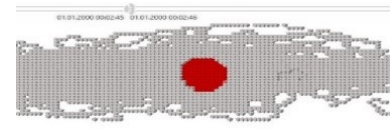
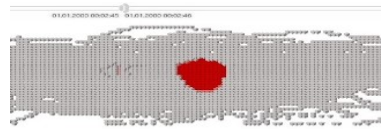


Abbildung 7-26: Heatmaps und Hotspots (links: Film mit räumlichem Sound, rechts: Film mit nicht räumlichem Sound), oben: Szene des Films, Mitte: Heatmaps, unten: signifikante Hotspots.

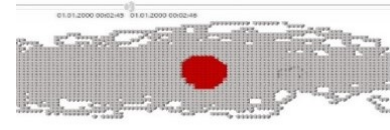
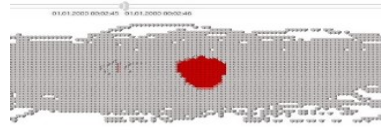
Wenn Leute in die Kamera sprechen, hat der Betrachter in CVR den Eindruck, dass die Person mit ihm spricht. In fast allen Fällen schauten die Betrachter mit nicht-räumlichem Sound ungefähr 8 s vor einem weiteren Umherschauen auf den Sprecher. Bei räumlichen Geräuschen dauerte es 1-3 s länger, bis nach weiteren Details gesucht wurde. In Abbildung 7-27 ist ein Beispiel dargestellt. Das Mädchen spricht zur Kamera (Betrachter) über ihr Leben. Die Szene beginnt bei Timecode 2:34. Bei 2:38 (*spatial*) und 2:39 (*nonspatial*) schauen fast alle Teilnehmer das Mädchen an. Nachdem das Mädchen angefangen hatte zu reden, änderten sich die Hotspots in beiden Gruppen 12 s lang nicht. Die Teilnehmer der *nonspatial*-Gruppe begannen jedoch früher, sich umzusehen, wie Abbildung 7-27 zeigt.



2:38 / 2:39



2:46



2:48

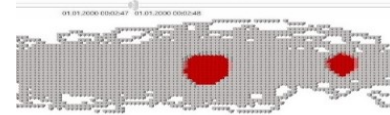
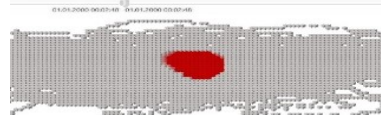


Abbildung 7-27: Szenenbild und signifikanter Hotspots für verschiedene Timecodes. Die zugehörigen Hotspots sind bis 12 s nach Sprechbeginn fast gleich.

Ähnliche Ergebnisse wurden für andere Timecodes im Film gefunden, wenn Personen zum Betrachter sprachen.

7.1.2.4 Diskussion und Einschränkungen

Die Muster der getrackten Kopfbewegungen beim Ansehen von CVR-Videos mit räumlichem und nicht räumlichem Sound waren für beide Gruppen meist sehr ähnlich. Zusammenfassend wurden die folgenden Ergebnisse gefunden:

- Sprechende oder aktive Personen lenken die Aufmerksamkeit des Betrachters mit räumlichem und nicht räumlichem Sound.
- Mit räumlichem Sound beobachten die Betrachter eine Person länger.
- Details ohne Geräusche ziehen bei nicht-räumlichen Sound mehr Aufmerksamkeit auf sich.

7.1.2.5 Fazit

In dieser Studie wurde untersucht, inwieweit sich das Blickverhalten beim Ansehen von Filmen mit räumlichem und nicht räumlichem Sound unterscheidet. Dabei wurde festgestellt, dass der Sound die Aufmerksamkeit des Betrachters auf sich zieht, auch wenn er nicht räumlich ist. Diese Ergebnisse können zur Integration von Cues in einen Film verwendet werden, um die Aufmerksamkeit des Zuschauers auf Dinge zu lenken, die für die Geschichte wichtig sind.

7.1.3 Studie: Guiding mittels PiP oder Fading

Die Studie ist Teil der Bachelorarbeit von Duc Anh Le (2017): Nondiegetic Methods for Guiding Attention in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Das Ziel dieser Studie war es, neue nicht-diegetische Methoden zur Lenkung der Aufmerksamkeit zu finden. Dafür wurden zwei Techniken implementiert und miteinander verglichen: PiP (Picture-in-Picture) und Fading (Aus-/Einblenden).

PiP

Die *PiP*-Methode blendet ein kleines Vorschauenfenster ein, welches die RoI (Region of Interest) zeigt. Das Fenster ist an der Seite des Displays zu sehen, welche näher an der RoI liegt (Abbildung 7-28, links). Der Zuschauer kann entscheiden, ob er seine Blickrichtung ändert.

fade

Die *fade*-Methode zeigt einen nach unten gerichteten Pfeil, wenn es eine neue RoI außerhalb des FoVs gibt (Abbildung 7-28, rechts). Der Zuschauer kann diesem Pfeil folgen, indem er durch sein Nicken eine Schwarzblende auslöst, nach der der Film neu ausgerichtet ist und der Zuschauer die neue RoI im FoV sieht.



Abbildung 7-28: Die *PiP*-Methode zeigt in einem Fenster die RoI (links). Ein Pfeil zeigt dem Nutzer an, dass er nach unten schauen kann, um einen Szenenwechsel auszulösen (rechts).

7.1.3.1 Material

Die Teilnehmer sahen sich sphärische Videos über eine Oculus Rift an. Für die Testumgebung und die Implementierung der Methoden wurde Unity 3D verwendet. Die beiden Videos zeigten jeweils Szenen aus dem Leben der Kelten, jedes aufgenommen im Inneren einer Hütte, in denen Protagonisten verschiedenen handwerklichen Tätigkeiten nachgingen. Die Protagonisten saßen und änderten nicht ihren Platz. Die Teilnehmer konnten die Ausführung der Tätigkeiten beobachten. Es gab keine weitere Story, um die Wirkung der Methoden nicht zu beeinflussen.

7.1.3.2 Parameter

In einer informellen Vorstudie (7 Teilnehmer) wurden die geeignetsten Parameter ermittelt. Für die *PiP*-Methode wurden verschiedene Größen des *PiP*-Fensters getestet. Zu große Fenster verdecken zu viel vom Filminhalt, auf zu kleinen Fenstern kann man den Bildinhalt nur schlecht erkennen. Ein Fenster von 15° x 25° erwies sich als geeignet und wurde für die Studie verwendet.

7.1.3.3 Teilnehmer und Prozedur

30 Personen (11 weiblich, 19 männlich, Alter: MW=24.3 Jahre) nahmen an der Studie teil. 21 von ihnen hatten keine oder wenig VR-Erfahrungen. Ein within-subject Testdesign wurde angewendet. Jeder Teilnehmer sah beide Videos, jedes mit einer anderen Methode. Dabei wurde sowohl die Reihenfolge der Videos als auch der Methoden permutiert.

Die folgenden abhängigen Variablen wurden für jede der beiden Methoden per Fragebogen untersucht: User Experience, Simulator-Sickness, Presence, Enjoyment, Recall. Der Fragebogen endete mit einem Vergleich der beiden Methoden.

7.1.3.4 Ergebnisse

User Experience, Simulator-Sickness, Presence, Enjoyment

Bei keinem dieser Faktoren konnten signifikante Unterschiede zwischen den beiden Methoden festgestellt werden.

Vergleich

70 % der Teilnehmer fanden die *PiP*-Methode **komfortabler**, 30 % die *fade*-Methode. Dabei wurden folgende Vor- und Nachteile für die jeweilige Methode angegeben:

Vorteile *PiP*:

- „man sieht die andere Region“ (P1, P4, P7, P16, P17, P18)
- „natürlicher“ (P5, P6, P22), „bessere Orientierung“ (P26, P27)
- „man kann selbst entscheiden, ob man hinschaut“ (P14)

Nachteile *PiP*:

- „ablenkender“ (P2, P8, P9, P11, P30)

Vorteile *fade*:

- „weniger ablenkend“ (P2, P19)
- „dezenter“ (P11, P12)
- „realistischer“ (P20), „natürlicher“ (P23)

Nachteile *fade*:

- „Teleportation“ (P1), „unphysiologisch“ (P6), „irritierend“ (P21)
- „man weiß nicht, wo man ist“ (P13)
- „nicht intuitiv“ (P14)
- „zu viel Aufwand, falls man nur mal kurz woanders schauen möchte“ (P14)

66.7 % der Teilnehmer fanden die *PiP*-Methode **hilfreicher**, 20 % die *fade*-Methode, für 13.3 % der Teilnehmer waren beide Methoden gleich hilfreich. Dabei wurden folgende Vor- und Nachteile für die jeweilige Methode angegeben:

Vorteile *PiP*:

- „man sieht die andere Region, bevor man sich entscheidet“ (P1,P3,P4,P6,P11,P14)
- „mehr Informationen“ (P10, P15, P18)
- „man verliert nicht den Pfad“ (P7)

Nachteile *PiP*:

- „Region schwerer zu finden“ (P24, P27)

Vorteile *Fade*:

- „man muss weniger wissen“ (P8), „leichter zu verstehen“ (P12, P24)
- „realistischer“ (P20)
- „man musste sich nicht selbst orientieren, um die Region zu finden“ (P29)

Nachteile *Fade*:

- „irritierend“ (P21)

7.1.3.5 Diskussion

Die *PiP*-Methode war für die meisten Teilnehmer sowohl komfortabler als auch hilfreicher. Der Vorteil dieser Methode liegt darin, dass man auf einen Blick erfassen kann, was die *RoI* beinhaltet und sich dann entscheiden kann, ob man sich in diese Richtung dreht. Allerdings muss die Region selbst gefunden werden, und das zusätzliche Fenster kann vom Inhalt des eigentlichen Sichtfeldes ablenken.

7.1.4 Studie: Guiding durch Flicker

Die Studie ist Teil der Masterarbeit von Ziheng Gao (2018): Gaze Direction by High Frequency Flickering in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Das Ziel dieser Studie war es herauszufinden, wie hilfreich oder störend ein Flicker ist, um den Betrachter zur Rol zu führen. Dafür soll die von Waldin beschriebene Methode (Waldin u. a., 2017) an CVR angepasst werden. Diese Methode nutzt aus, dass ein in der Peripherie flackerndes Signal im Bereich der Fovea nicht als flackernd wahrgenommen wird. Der Flicker fusioniert zu einem stabilen Bild. Die Flicker-Methode wurde sowohl mit einem Pfeil wie auch mit einer Baseline verglichen:

Flicker

Ist die Rol nicht im Blickfeld des Betrachters, wird ein schmaler flickernder Streifen an der Seite des Displays erzeugt, die näher an der Rol ist (screen-referenziert). Sobald sich die Rol im FoV befindet, wird ein kleiner runder Flicker direkt auf die Rol platziert (world-referenziert).

Arrow

Ein Pfeil zeigt in Richtung der Rol, unabhängig davon, ob diese bereits im FoV zu sehen ist oder nicht.

Baseline

Der Film wird ohne Methoden zur Aufmerksamkeitslenkung angeschaut.

7.1.4.1 Material

Die Teilnehmer sahen sich die Filmsequenzen über eine Oculus Rift an. Für die Testumgebung und die Implementierung des Flickers wurde Unity 3D verwendet. Für diese Studie wurde ein 3-minütiges CVR-Video des Crossing Border Projektes genutzt (Vtopia360, 2017).

7.1.4.2 Parameter

In den ursprünglichen Untersuchungen von Waldin (Waldin u. a., 2017) wurden Flickerfrequenzen von 60 Hz und 72 Hz verwendet. Wir konnten dies nicht übernehmen, da die Frequenz des Displays (Oculus Rift) 90 Hz beträgt und daher die höchstmögliche Flickerfrequenz bei 45 Hz liegt. Diese Frequenz wurde in unserem Experiment genutzt. Das FoV der Oculus Rift beträgt ca. 90°, so dass der screen-referenzierte Flicker in einem Winkel von 45° seitlich der Kopfrichtung erschien.

7.1.4.3 Teilnehmer und Prozedur

24 Personen (10 weiblich, 14 männlich, Alter: MW= 24.7 SD=2.1) nahmen an der Studie teil. 75 % von ihnen hatten keine oder wenig VR-Erfahrungen. Die Teilnehmer wurden in zwei Gruppen aufgeteilt und sahen sich das Video zweimal an, jede Gruppe mit anderen Lenkungsmethoden. Die Gruppe A mit Flickern und Pfeil, die Gruppe B mit Flickern und eine Baseline ohne Lenkung. Die Reihenfolge der Methoden wurde permutiert.

Die folgenden abhängigen Variablen wurden per Fragebogen ermittelt: Enjoyment, Benutzererfahrung (UEQ), Presence (PQ), Recall. Der Fragebogen endete mit einem Vergleich der beiden Methoden.

Enjoyment: auf einer 7-Punkte-Likert-Skala wurden die folgenden Fragen beantwortet:

(E1) Wie subtil ist die Methode?

(E2) Hat die visuelle Szene Ihre Aufmerksamkeit erregt?

- (E3) Hat die Methode Ihre Aufmerksamkeit erregt?
 (E4) Hat die Methode Ihre Aufmerksamkeit von der Szene abgelenkt?
 (E5) Hat Ihnen die CVR-Experience gefallen?
 (E6) Möchten Sie die Applikation wiederholen?

Für Frage (E1) reichte die Skala von „subtil“ (1) bis „offensichtlich“ (7), für die Fragen (E2) bis (E6) von „überhaupt nicht“ (1) bis „sehr“ (7).

Benutzererfahrung: Um herauszufinden, wie die Flickermethode die Benutzererfahrung beeinflusst, wurden Items des User Experience Questionnaires UEQ genutzt (Laugwitz u. a., 2008) und die Bewertungen der Flickermethode mit denen der Baseline und der Pfeilmethode verglichen.

Für die Recall-Aufgabe wurden 4 Fragen zu visuellen Details im Film gestellt.

7.1.4.4 Ergebnisse

Enjoyment

Flicker-Arrow: Die Teilnehmer empfanden den Flicker subtiler als den Pfeil ($MW_{\text{Flicker}} = 4.17$ $SD_{\text{Flicker}} = 1.69$, $MW_{\text{Pfeil}} = 5.6$ $SD_{\text{Pfeil}} = 1.5$, $p_{\text{t-test}} = 0.02$). Zusätzlich lenkte der Flicker die Aufmerksamkeit weniger von der Szene ab ($MW_{\text{Flicker}} = 3.79$, $SD_{\text{Flicker}} = 1.64$, $MW_{\text{Pfeil}} = 5.13$, $SD_{\text{Pfeil}} = 1.64$, $p_{\text{t-test}} = 0.3$). Abbildung 7-29 zeigt die Mittelwerte für alle Items.

Flicker-Baseline: Der Enjoyment-Teil des Fragebogens enthielt drei Fragen zu den Lenkungsmethoden (E1, E3, E4), die nicht für die Baseline gestellt wurden. Bei den anderen Items (E2, E5, E6) war der Unterschied zwischen Flicker-Methode und Baseline nicht signifikant ($p_{\text{t-test}} > 0.1$).

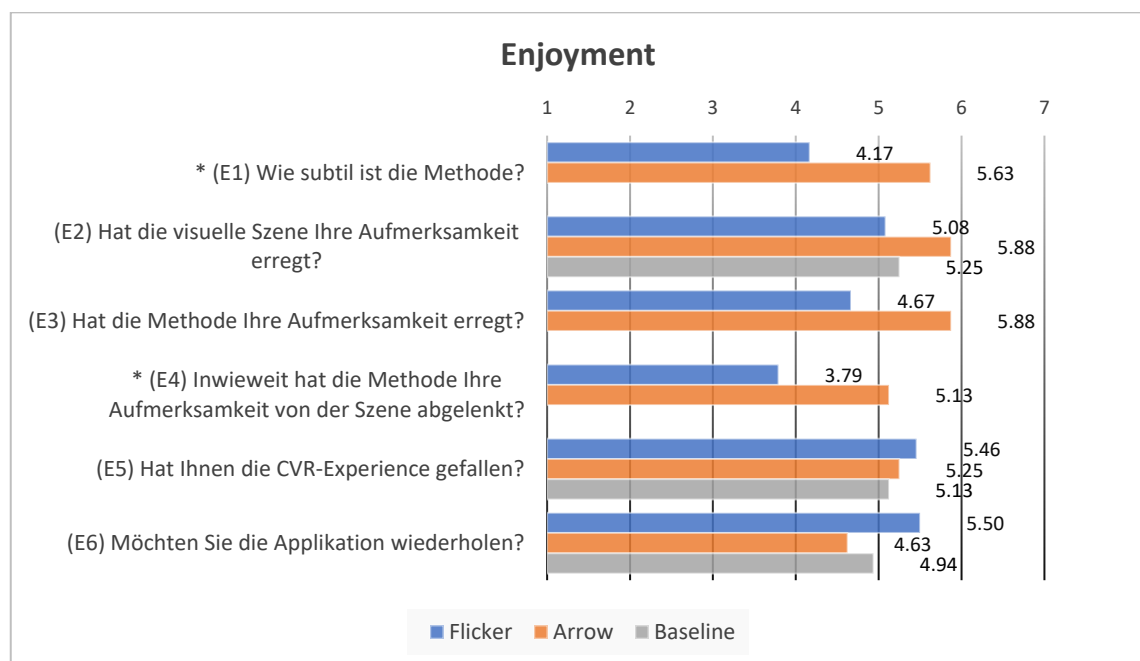


Abbildung 7-29: Mittelwert für die Enjoyment-Items (Likert-Skala von 1 bis 7). Signifikante Unterschiede sind mit * gekennzeichnet.

User Experience

Beim Vergleich der Benutzererfahrung (Abbildung 7-30) konnten die folgenden signifikanten Unterschiede festgestellt werden:

Flicker-Arrow:

- angenehm: Der Flicker ist angenehmer als der Pfeil
 ($MW_{\text{Flicker}} = 5.5$ $SD_{\text{Flicker}} = 1.14$, $MW_{\text{Pfeil}} = 4.63$ $SD_{\text{Pfeil}} = 1.6$ $p = 0.05$)

- unterstützend: Der Flicker ist unterstützender als der Pfeil
($MW_{\text{Flicker}}=5.29$ $SD_{\text{Flicker}}=0.86$, $MW_{\text{Pfeil}}=4.63$ $SD_{\text{Pfeil}}=0.9$ $p = 0.05$)

Flicker-Baseline:

- unterstützend: Der Flicker ist unterstützender als die Baseline
($MW_{\text{Flicker}}=5.29$ $SD_{\text{Flicker}}=0.86$, $MW_{\text{Base}}=4.56$ $SD_{\text{Base}}=0.8$ $p_{\text{t-test}} = 0.01$)
- innovativ: Der Flicker ist innovativer als die Baseline
- ($MW_{\text{Flicker}}=5.33$ $SD_{\text{Flicker}}=0.87$ $MW_{\text{Base}}=4.75$ $SD_{\text{Base}}=0.9$ $p_{\text{t-test}} = 0.02$)

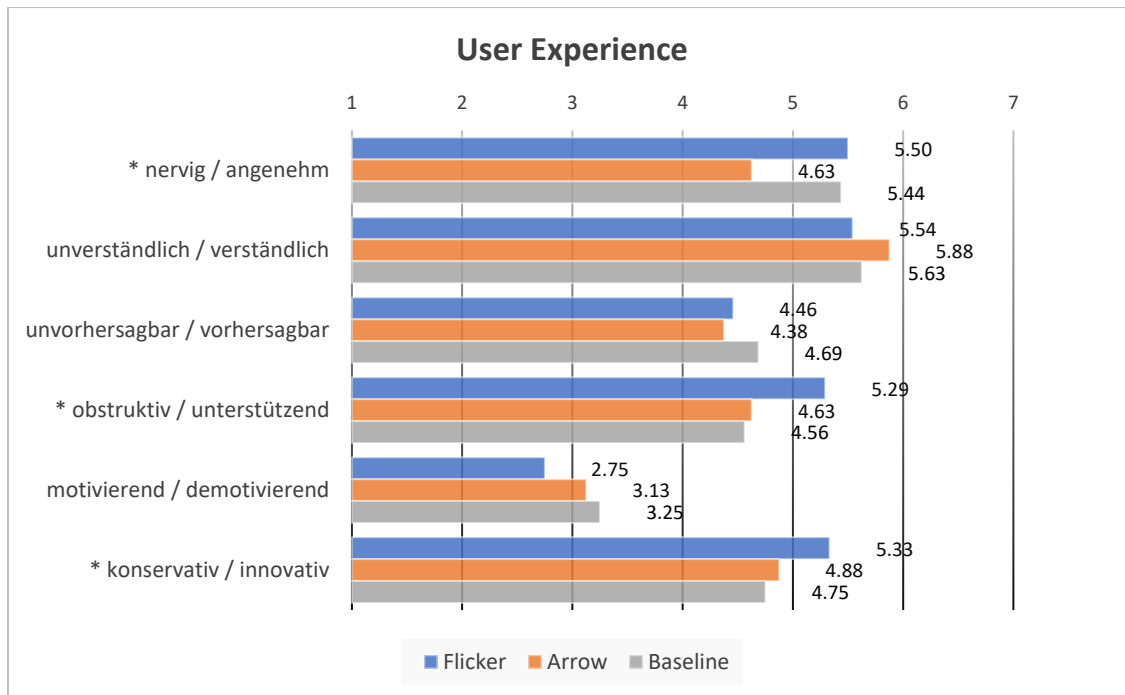


Abbildung 7-30: Elemente des UEQs (Likert-Skala von 1 bis 7). Das Flickern war angenehmer als der Pfeil und hilfreicher als Pfeil und Baseline.

Presence

Flicker-Arrow: Bei der Benutzung des Pfeils war die Presence niedriger als bei Nutzung der Flicker-Methode: $MW_{\text{Flicker}}=5.17$ $SD_{\text{Flicker}}=0.96$, $MW_{\text{Pfeil}}=4.25$ $SD_{\text{Pfeil}}=0.9$ $p_{\text{t-test}} = 0.03$.

Flicker-Baseline: Bezüglich der Presence-Items gab es keine signifikanten Unterschiede zwischen der Baseline und der Flicker-Methode.

Recall

Die Recall-Rate war für die Flicker-Methode höher als für die beiden anderen Methoden. Bei der Flicker-Methode waren 47.5% der Antworten richtig, bei der Baseline 37.5% und bei der Arrow-Methode 20%.

Vergleich

Flicker-Arrow: 75 % der Teilnehmer bevorzugten den Flicker: „Der Flicker beeinträchtigt nicht die Immersion“ / „ist nicht so offensichtlich“ (P1, P2, P4, P5, P8) und die „Pfeile sind etwas störend“ (P8). 25 % der Teilnehmer bevorzugten den Pfeil: „Das Flickern stört die Augen“ (P4) oder „Eindruck eines technischen Problems“ (P6),

Für die Hälfte der Teilnehmer der Gruppe A war die Flicker-Methode hilfreicher als der Pfeil, für 25 % der Pfeil. Das „Flickern lenkt nicht vom Genießen des Videos ab“ (P4, P5, P8) und der Pfeil „war leicht zu verstehen“ (P3, P6) und „zu erkennen“ (P1). Die anderen 25% hatten keine Präferenz.

Flicker-Baseline: Vergleicht man das Flickern mit der Baseline, fanden es 87.5 % der Teilnehmer hilfreich, gelenkt zu werden. Es „unterstützt die Konzentration“ (P9), hilft, „die richtigen Stellen zu finden“ (P11, P16, P17, P19, P20, P21, P22, P24). Für andere Teilnehmer war die Lenkung „etwas verwirrend“ (P14), sie zogen es vor, „die Kontrolle zu übernehmen und keine Anweisungen zu befolgen“ (P13).

In Gruppe B bevorzugte die Hälfte der Teilnehmer, sich vom Flickern leiten zu lassen und die andere Hälfte, das Video ohne Lenkung anzusehen. Das Flickern „half, den Fokus zu bekommen“ (P10, P19, P21, P24), aber es „unterbrach das visuelle Erlebnis“ (P9, P14, P22), und auch hier empfanden einige Teilnehmer den Flicker als ein technisches Problem der Anwendung (P13, P23).

50 % der Zuschauer fühlten sich mit der Lenkungsmethode nicht wohl, 87.5 % hielten es jedoch für hilfreich, sich ein CVR-Video damit anzusehen. Teilnehmer, denen die Flicker-Methode nicht zusagte, gaben an, dass diese das Seherlebnis „unterbricht“ (P1, P6, P14) oder „etwas schwindlig macht“ (P3, P10). Die Teilnehmer, die mit der Methode zufrieden waren, gaben an, „dass es hilfreich ist, die richtige Blickrichtung zu finden“ (P2, P11, P13, P16) und „einfach zu verstehen“ ist (P4).

7.1.4.5 Diskussion

In den drei Faktoren Enjoyment, Benutzererfahrung und Presence konnten keine signifikanten Unterschiede zwischen der Flicker-Methode und der Baseline festgestellt werden. Darüber hinaus wurde ein positiver Effekt auf die Recall-Rate für Details erfasst, zu denen der Betrachter gelenkt wurde. Dies ist konsistent zu den Ergebnissen von Veas et al. (Veas u. a., 2011), dass SMT-Guiding Methoden die Recall-Rate erhöhen.

Die Flicker-Methode erwies sich als wenig störend und minderte das Erlebnis im Vergleich zur Baseline nicht. Die weniger subtile Arrow-Methode beeinträchtigte jedoch das Erlebnis und wirkte störender.

Allerdings war der Flicker nicht bei allen Teilnehmern wirklich subtil und nicht alle Teilnehmer empfanden den Flicker als angenehm. Einige waren irritiert und interpretierten ihn als technische Störung. Trotzdem fanden es die meisten Teilnehmer hilfreich, sich lenken zu lassen, da es einfacher ist, wichtige Details nicht zu übersehen. Dabei wurde der Flicker dem Pfeil vorgezogen, weil er das Erlebnis weniger stört.

7.1.5 Studie: Guiding durch subtile Flicker und Eyetracking

Die Studie ist Teil der Bachelorarbeit von Felix Althammer (2018): Guiding the User's Gaze in 360 Degree Videos.

Diese Arbeit wurde von mir gemeinsam mit Mohamed Khamis betreut. Die gemeinsam entwickelte Idee implementierte Felix Althammer, wobei die Projektschritte in regelmäßigen Treffen besprochen wurden. Der Student führte die Studie in Abstimmung mit Mohamed Khamis und mir durch. Die Daten habe ich neu analysiert und mit Felix Althammer und Mohamed Khamis diskutiert.

Die Ergebnisse dieser Studie wurden vorab veröffentlicht in:

- Rothe, S., Althammer, F., Khamis, M., 2018. GazeRecall: Using Gaze Direction to Increase Recall of Details in Cinematic Virtual Reality, in: Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia - MUM '18. ACM

In dieser Studie wurde untersucht, ob SGD (Subtil Gaze Direction) (Bailey u. a., 2009) mittels Eyetracking für CVR anwendbar ist. Dabei wurde zusätzlich überprüft, wie gut Benutzer sich an Details erinnern können, auf die sie gelenkt wurden. Dies ist wichtig, um zu wissen, ob die Blicklenkung auch dazu führt, dass wichtige Details wahrgenommen werden oder nur die Blickrichtung angepasst wird.

Da die Wahrnehmung von Flickern von vielen personenbezogenen Kenngrößen abhängen (Abschnitt 2.1.2) war es nicht möglich, Flickerparameter zu finden, die für alle Personen sowohl subtil als auch wirksam sind. Wir entschieden uns deshalb zwei Fälle miteinander zu vergleichen:

A: Die Intensität des Flickers liegt für die meisten Teilnehmer an der Grenze zwischen subtil und wahrnehmbar.

B: Die Intensität des Flickers ist so, dass er für alle Personen deutlich wahrnehmbar ist.

Wie in der Original SGD-Methode von Bailey (Bailey u. a., 2009) wurde der RoI ein 10-Hz-Flicker hinzugefügt, der verschwindet, wenn der Betrachter darauf schaut. Allerdings benötigt man in CVR eine zusätzliche off-screen Methode: befindet sich die RoI nicht im Sichtfeld, erscheint am Rand des Displays ein Flicker, je nachdem, ob die RoI sich näher am linken oder rechten Rand befindet. Ein Toleranzbereich wurde hinzugefügt, um zu verhindern, dass der Flicker zwischen links und rechts springt, wenn sich die RoI hinter dem Betrachter befindet.

7.1.5.1 Material

Es wurde eine HTC Vive mit integriertem Pupil Labs Eyetracker verwendet. Die Teilnehmer sahen sich ein Video an (Auflösung 4096x2048, 25fps), das aus drei Szenen besteht. Die erste zeigt ein keltisches Dorf, in dem ein Druide etwas über das Leben der Kelten erzählt. In der zweiten Szene sitzt der Druide in einem keltischen Haus und erklärt einige Details des Alltags. Danach ist der Raum mit keltischen Dorfbewohnern gefüllt, die verschiedene handwerkliche Tätigkeiten ausüben (Abbildung 7-31). Die ersten beiden Szenen dienten der Eingewöhnung, in der dritten Szene wurden drei Objekte als Ziel der Aufmerksamkeitslenkung ausgewählt: eine Frau an einem Webstuhl, ein Mann mit Bart und ein Speer. Da der Betrachter sich zu Beginn einer neuen Szene umsehen möchte und deshalb eine Aufmerksamkeitslenkung schwierig ist (**Studie 7.1.1**), setzte das Flickern erst nach 15 s ein. Nach 20 s wurde das nächste Ziel mit einem Flicker versehen. Die Reihenfolge der verschiedenen Ziele wurde mit einem lateinischen Quadrat ausbalanciert.



Abbildung 7-31: Verwendete Szene: die Bereiche, zu denen der Betrachter geführt wurde, sind rot markiert.

7.1.5.2 Parameter

Die Frequenz und die Größe des Flickers wurden der ursprünglichen Methode von Bailey (Bailey u. a., 2009) entnommen: 10 Hz, 0.76°. In einer informellen Vorstudie wurden zusätzlich verschiedene Farbtintensitäten getestet, um herauszufinden, welche Werte subtil und dennoch wirksam sind. Die *Farbtintensität* ist ein Wert im Intervall [0,1], der die Intensität der Modulationen bestimmt. In der Vorstudie sahen sich fünf Teilnehmer ein kurzes Video an und wurden angewiesen, einen gelben Kreis in der Mitte des FoV zu betrachten. Für einen Flicker in der Peripherie wurde die Intensität um 0.1 Stufen erhöht, bis der Teilnehmer ihn bemerkte. Für jeden Teilnehmer wurden die gleichen 8 Orte für den Flicker ausgewählt, mit unterschiedlichen Hintergründen von dunkel bis hell. Zusätzlich wurde nach einem Intensitätswert gesucht, der offensichtlich war und als obere Schwelle in der Hauptstudie verwendet werden konnte. Dafür wurde die Intensität des Flickers so lange erhöht, bis der Teilnehmer keine Änderung des Flickers mehr erkennen konnte.

Die Vorstudie zeigte folgenden Ergebnisse: (1) Wird ein sehr heller Punkt betrachtet, ist es schwierig in der Nähe einen Flicker wahrzunehmen. (2) Veränderungen des Sehvermögens können sich aus einer Anpassung der Pupille an helle Szenen ergeben und sich auf die Wahrnehmung von Dingen an der Peripherie auswirken. Dies bedeutet, dass die Attribute des Flickers angepasst werden müssen, je nachdem, wo der Benutzer hinschaut.

7.1.5.3 Teilnehmer und Prozedur

Wir entschieden uns für ein between-subject Testdesign und teilten die 42 Teilnehmer nach dem Zufallsprinzip in zwei Gruppen ein: Gruppe A (5 weiblich, 16 männlich, Alter: MW=24.9Jahre), Gruppe B (9 weiblich, 12 männlich, Alter: MW=25.4 Jahre). Alle Teilnehmer sahen sich das gleiche Video an, jede Gruppe mit einer anderen Flickerintensität.

Für die Gruppe A wurde eine geringere Flickerintensität verwendet (0.2 für weiß und 0.5 für schwarz). Gruppe B schaute sich das gleiche Video mit einem deutlicheren Flicker (Intensität = 1) an. Der Film besteht aus drei Szenen, wobei die ersten beiden Szenen der Eingewöhnung dienten.

Die Teilnehmer sahen sich die Videos in stehender Haltung an und konnten Kopf und Körper drehen. Es gab keine spezielle Aufgabe, sie sollten sich umsehen und den Objekten folgen, an denen sie interessiert sind. Nach dem Ansehen des Videos füllten die Teilnehmer einen Teil des Fragebogens aus.

Die folgenden abhängigen Variablen wurden per Fragebogen ermittelt: Enjoyment, Presence (PQ) und Recall. Es wurden ähnliche Fragen wie in **Studie 7.1.4** verwendet, allerdings in einer gekürzten Form da wir uns auf die Frage konzentrierten, ob die Flickerintensität die Recall-Rate beeinflusst. Zusätzlich zeichneten wir Kopf- und Augenbewegungen auf, um zu analysieren, wohin die Betrachter schauen.

Alle Fragen mit Ausnahme des Recall-Teils wurden über 7-Punkte-Likert-Skalen beantwortet und für jedes Item wurde ein t-Test durchgeführt. Die Teilnehmer wurden zu drei verschiedenen Rols gelenkt und nach visuellen Details in diesen Bereichen gefragt (Recall).

7.1.5.4 Ergebnisse

Enjoyment

Das Enjoyment („Wie hat Ihnen die Anwendung gefallen?“) war in Gruppe A signifikant höher als in Gruppe B:

- $MW_A = 5.19$, $SD_A = 1.36$
- $MW_B = 4.43$, $SD_B = 1.4$
- $p_{t\text{-test}} = 0.04$

Presence

Der t-Test zeigte keine signifikanten Unterschiede auf (p-Werte zwischen 0.16 und 0.5).

Recall

Um zu ermitteln, ob Details auf Grund der Methode unterschiedlich aufgenommen werden, wurden drei Fragen zu Details von Objekten in den Rols gestellt: die Farbe der Schürze der Frau am Webstuhl, die Form des Bartes des Mannes und das Vorhandensein des Speeres. Die Recall Rate war für Gruppe B besser (Abbildung 7-32). Die Form des Bartes und die Farbe der Schürze konnten von mehr Leuten reproduziert werden, wenn der offensichtlichere Flicker verwendet wurde. Die Teilnehmer nahmen der Speer nicht wahr, was durch die Position (höher als die beiden anderen) oder den hellen Bereich (die offene Tür) in der Nähe verursacht sein könnte.

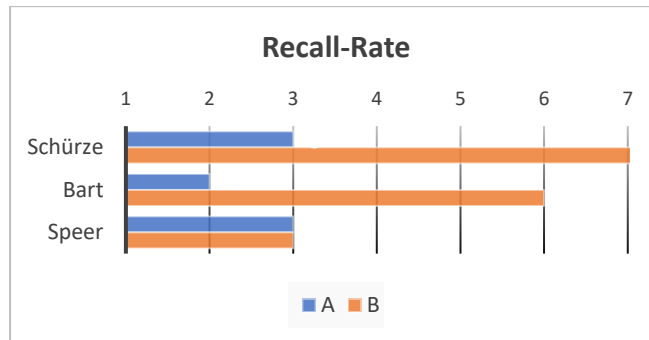


Abbildung 7-32: Recall Rate für die drei Objekte in Prozent.

Der exakte Fisher-Test zeigte einen signifikanten Unterschied zwischen den Recall-Raten von Gruppe A und B ($p = 0.01$).

Analyse der Tracking-Daten

Bei der Analyse der Heatmaps für Kopf- und Eyetracking-Daten konnte festgestellt werden, dass Bewegungen im Video den Blick stärker lenkte als das Flickern (Abbildung 7-33, Abbildung 7-34).

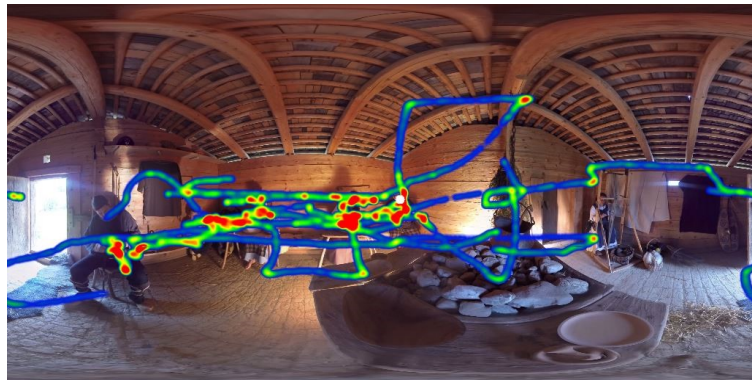


Abbildung 7-33: Heatmap der Headtracking-Daten. Der Flicker ist in der Mitte des Bildes. Die Daten bilden Hotspots, um bewegte Dinge - nicht nur in der Nähe des Flickers.

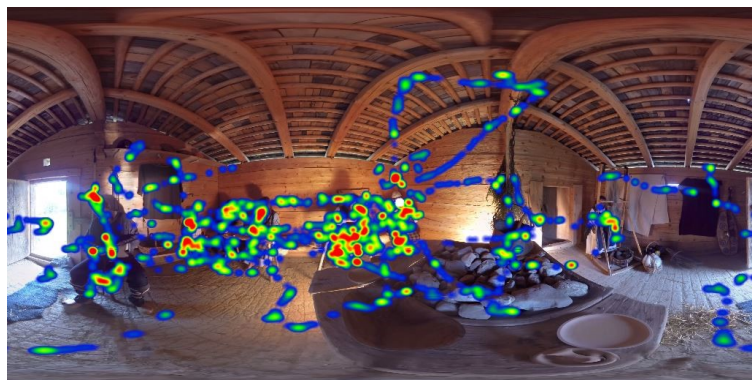


Abbildung 7-34: Heatmap der Eyetracking-Daten, die zu den Headtracking-Daten von Abbildung 7-33 gehören.

Wie im Recall-Absatz bereits erwähnt, war der Speer in der Nähe der Tür nicht im Fokus der Teilnehmer, was auch an der Heatmap erkennbar war (Abbildung 7-35).

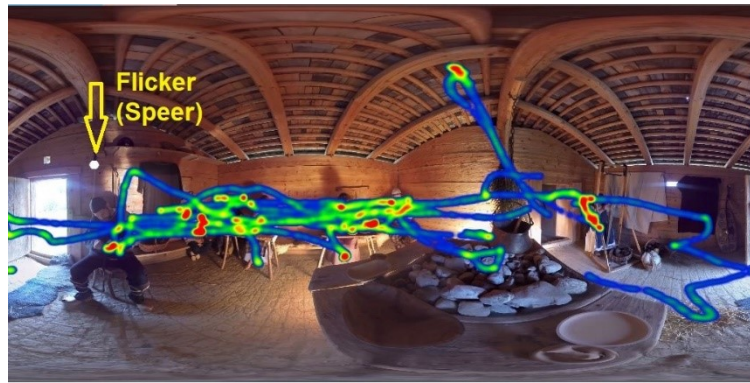


Abbildung 7-35: Der Flicker befindet sich auf dem Speer auf der linken Seite des Bildes (weißer Punkt rechts oben neben der Tür). Daten bilden Hotspots um sich bewegende Objekte, nicht in der Nähe des Flickers.

7.1.5.5 Diskussion

In unseren Experimenten wirkte der subtile Flicker nicht stärker als Bewegungen oder andere interessante Dinge im Film. Auch wenn eine subtile Lenkung in Standbildern geeignet ist (Waldin u. a., 2017), konkurrieren sie in bewegten Bildern mit Cues im Film, die die Aufmerksamkeit des Betrachters auf sich ziehen. Das Flickern mit der höheren Intensität, das nicht subtil war, konnte den Benutzer jedoch in den meisten Fällen lenken. Die höhere Flickerintensität erhöhte die Recall-Rate, verringerte jedoch das Enjoyment. Daher können Methoden mit starken Flickern in Fällen eingesetzt werden, in denen das Enjoyment nicht im Vordergrund steht, sondern das Aufnehmen und Merken von Details.

7.1.6 Studie: Guiding durch hochfrequente Flicker

Die Studie ist Teil der Masterarbeit von Ziheng Gao (2018): Gaze Direction by High Frequency Flickering in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und mit dem Studenten gemeinsam die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Die Aufgabe dieser Anwenderstudie bestand darin, Varianten der SGD-Methode zu untersuchen, für die kein Eyetracker erforderlich ist. Dem Target wird dafür ein hochfrequenter Flicker hinzugefügt (Waldin u. a., 2017). Erscheint dieser in der Peripherie wird er flickernd wahrgenommen und sollte basierend auf der Augenphysiologie verschwinden, wenn der Betrachter darauf schaut. In CVR benötigt man zusätzliche off-screen Lenkung, für den Fall, dass sich die RoI und damit der Flicker nicht im FoV des Betrachters befindet. Dazu werden in dieser Studie zwei Methoden verwendet:

discrete

Am rechten, bzw. linken Displayrand erscheint ein schmaler Flickerstreifen, je nachdem in welcher Richtung sich die RoI befindet. Diese Methode ist screen-referenziert.

continous

Ein runder Flicker erscheint so lange, bis der Zuschauer in seine Richtung schaut. Ist danach die RoI noch nicht im FoV erscheint ein weiterer Flicker. Diese Methode ist world-referenziert.

7.1.6.1 Material

Die Teilnehmer sahen sich die CVR-Applikation mit einer Oculus Rift an, die mit einem Desktop-Computer verbunden war. Die Flicker-Methoden wurden in Unity 3D implementiert. Zwei Videos der **Studie 7.1.5** wurden wiederverwendet (Szene 1 und 2).

7.1.6.2 Parameter

In den Untersuchungen von Waldin wurde eine Frequenz von 60 Hz und 72 Hz genutzt. Wir konnten dies nicht übernehmen, da die Frequenz unseres Displays (Oculus Rift) 90 Hz beträgt, wodurch sich eine höchstmögliche Flickerfrequenz von 45 Hz ergibt, die in unseren Experimenten verwendet wurde.

Um die Parameter für die Hauptstudie festzulegen, führten wir eine informelle Vorstudie mit 6 Teilnehmern durch. Dazu wurden statische 360°-Bilder und eine Flickerfrequenz von 45 Hz genutzt. Größe, Farbe und Luminanz wurden geändert. Die Vorstudie ergab, eine geeignete Flickergröße von 7.5°. Die Wirkung von Farbe und Luminanz des Flickers wird vom Hintergrund mit beeinflusst. Da die CFF stark vom Betrachter abhängt (Abschnitt 2.1.2), blendeten wir den Flicker in der Hauptstudie aus, wenn der Kopf des Betrachters in die Richtung der Rol zeigte.

7.1.6.3 Teilnehmer und Prozedur

16 Personen (10 männlich, 6 weiblich, Alter: $MW=26.1/SD=2.7$) nahmen an der Studie teil. 75 % hatten keine oder wenig Erfahrungen in VR. Es wurde ein within-subject Testdesign verwendet: Alle Teilnehmer sahen zwei Videos, jedes Video mit einer der beiden Methoden. Die Reihenfolge von Video und Methode wurde ausbalanciert. Alle Kopfbewegungen wurden aufgezeichnet und die folgenden abhängigen Variablen mittels Fragebogen ermittelt: Enjoyment, Task Workload (NASA-TLX), Simulator-Sickness (SSQ), Presence (PQ). Der Fragebogen endete mit Fragen zum Vergleich der beiden Methoden.

7.1.6.4 Ergebnisse

Es gab mehrere signifikante Unterschiede zwischen den beiden Methoden. Bezüglich der Presence unterschieden sich die Methoden nur für das Item „Ich fühlte mich im virtuellen Raum präsent.“, bei der die *continous*-Methode besser abschnitt: $MW_{cont} = 5.44$, $SD_{cont} = 0.81$, $MW_{discr} = 4.94$, $SD_{discr} = 0.68$, $p = 0.03$ (t-Test).

Beim Vergleich der SSQ-Scores ergaben sich für zwei Items signifikante Unterschiede, beide waren bei der *discrete*-Methode höher (Abbildung 7-36):

- Kopfschmerzen $MW_{discr} = 0.31$, $SD_{discr} = 0.6$, $MW_{cont} = 0.06$, $SD_{cont} = 0.25$, $p = 0.07$ (t-Test)
- Fokussierungsschwierigkeiten $MW_{discr} = 0.94$, $SD_{discr} = 0.77$, $MW_{cont} = 0.5$, $SD_{cont} = 0.5$, $p = 0.03$ (t-Test).

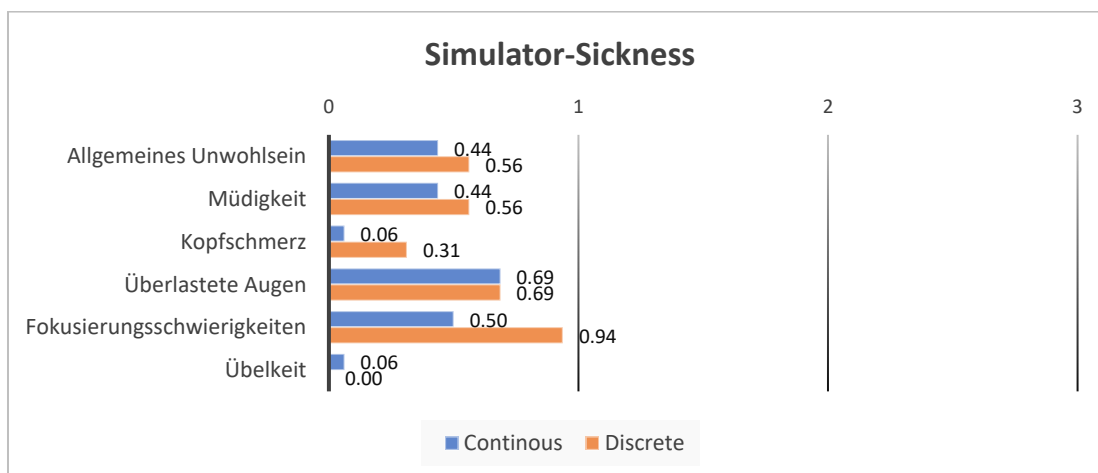


Abbildung 7-36: Mittelwerte für die SSQ-Scores (Likert-Skala von 0 bis 3).

Bei den Enjoyment-Items gab es keinen signifikanten Unterschied. Die Frage, welche Methode bequemer sei, lieferte für beide Methoden die gleichen Scores. 68.75 % fanden die kontinuierliche Methode jedoch hilfreicher. Teilnehmer, die die *continuous*-Methode bevorzugten, gaben an, dass sie „weniger ablenkend“ (P1, P5, P9), „angenehmer“ (P14, P15) und „leichter für die Konzentration“ (P4, P10) sei.

7.1.6.5 Diskussion

Für off-screen Guiding schnitt in unserer Studie die kontinuierliche screen-referenzierte Methode (*continuous*-Methode) besser ab. Die meisten Teilnehmer fanden diese Methode hilfreicher. Darüber hinaus war es bei der *discrete*-Methode schwieriger, sich zu konzentrieren.

Die Wirksamkeit von Flicker-Methoden hängt vom verwendeten Videomaterial und der verwendeten Hardware ab. Die Helligkeit des Bild-Hintergrundes hat einen großen Einfluss auf die Wirksamkeit des Flickers. In unserer Studie haben wir für alle Hintergründe dieselben Parameter verwendet. In zukünftiger Forschung sollte untersucht werden, ob an den Hintergrund angepasste Flicker effektiver wirken können. Die Auflösung und Frequenz der heutigen Displays schränken die Wirksamkeit subtiler Flicker-Methoden ein. Methoden wie SGD mit hochfrequenten Flickern (Waldin u. a., 2017) erfordern eine Displayfrequenz von mehr als 90 Hz und ein breiteres Sichtfeld.

7.2 Nutzerstudien: Kameraführung

Um die geeignete Position der Kamera zu finden, müssen verschiedene Dinge berücksichtigt werden. Zum einen die Abstände und räumlichen Relationen zwischen Charakteren und Kamera (**Studie 7.2.1** und **7.2.2**), die Höhe der Kamera (**Studien 7.2.3** und **7.2.4**) und mögliche Kamerabewegungen (**Studie 7.2.5**). Die Ergebnisse dieser Studien werden in Kapitel 3 erörtert und eingeordnet.

7.2.1 Studie: Proxemiks und Kameraabstände

Die Studie ist Teil der Bachelorarbeit von Pia Probst (2020): Proxemics in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit der Studentin die Idee entwickelt, die sie implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir von der Studentin durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Ziel dieser Studie war es, herauszufinden, in welcher Art verschiedene Kameraabstände auf den Zuschauer wirken. Dazu wurden Distanzen der Proxemik (Hall, 1966) genutzt. Die Kamera (Sicht des Zuschauers) nahm die folgenden Abstände von einer zur Kamera sprechenden Person ein (Abbildung 7-37):

intimate: 0.45 m Abstand zwischen Kamera und Sprecherin (intim, Abk.: *int*)

personal: 1 m Abstand zwischen Kamera und Sprecherin (persönlich, Abk.: *pers*)

social: 2.5 m Abstand zwischen Kamera und Sprecherin (sozial, Abk.: *soc*)

public: 5 m Abstand zwischen Kamera und Sprecherin (öffentlich, Abk.: *pub*)



Abbildung 7-37: Sprechende Person in den Videosequenzen in verschiedenen Abständen, von links nach rechts: intim, persönlich, sozial, öffentlich. Die Bilder sind in dieser Abbildung verfremdet.

7.2.1.1 Material

Für diese Studie wurden 17 Videosequenzen im Musikautomatenraum des Deutschen Museums produziert. Eine Szene diente als Einstiegsszene, um dem Nutzer zu ermöglichen, sich mit der Technik vertraut zu machen und sich im Raum umzusehen. Alle anderen Szenen zeigten eine Sprecherin, die Informationen zu den ausgestellten Exponaten vermittelte. Dafür wurden von der Sprecherin an vier verschiedenen Kamerastandpunkten je 4 vier verschiedene Distanzen zur Kamera eingenommen. Der Sound war räumlich, so dass die Stimme aus der Richtung der sprechenden Person kam. Während der Studie nutzten die Teilnehmer eine HTC Vive und Kopfhörer.

7.2.1.2 Teilnehmer und Prozedur

Es nahmen 24 Personen (12 weiblich, 12 männlich, Alter: $MW=23$, $SD=5.6$) an der Studie teil. 16 von ihnen hatten bereits VR-Erfahrungen, vier von ihnen nutzten regelmäßig ein HMD.

Es wurde ein within-subject Testdesign verwendet, und jeder Teilnehmer sah die Filmsequenzen mit allen vier Abständen. Die Reihenfolge der Abstände wurde durch Permutationen ausbalanciert. Zu Beginn der Studie sah jeder Teilnehmer ein Einführungsvideo, welches den Raum der Aufnahmen zeigte, jedoch ohne die Sprecherin. Die Teilnehmer wurden aufgefordert sich umzuschauen, um den Novelty-Effekt und die intensivere Inspektion der Umgebung im ersten Teil der Studie zu reduzieren. Danach beantworteten die Teilnehmer demographische Fragen, nach jeder Videosequenz Fragen zum eben gesehenen und am Ende der Studie zusätzliche Fragen zum Vergleich der Videos.

Es wurden Teile des Presence-Fragebogens (Schubert u. a., 2002), des SSQ (Kennedy u. a., 1993) und des UEQ (Hinderks u. a., 2018) genutzt. Zusätzlich wurden Fragen zu den Empfindungen während des Videos gestellt sowie Head- und Eyetracking-Daten aufgezeichnet.

7.2.1.3 Ergebnisse

Presence und allgemeines Befinden

Bezüglich des Presence-Items „In der computererzeugten Welt hatte ich den Eindruck, dort gewesen zu sein“ erzielte die *pub*-Bedingung signifikant niedrigere Scores als die *pers*-Bedingung: $MW_{pers}=4.5$ $SD_{pers}=1.06$ $MW_{pub}=3.96$ $SD_{pub}=1.23$ $p_{pers/pub}=0.04$.

Im int-Fall empfanden wesentlich mehr Teilnehmer allgemeines Unbehagen ($MW_{int}=1.74$ $SD_{int}=1.71$ $MW_{pers}=0.79$ $SD_{pers}=1.5$ $MW_{soc}=0.83$ $SD_{soc}=1.4$ $MW_{pub}=0.92$ $SD_{pub}=1.14$) und fühlten sich irritiert und gestresster als unter den anderen drei Bedingungen ($MW_{int}=2.17$ $SD_{int}=2.06$ $MW_{pers}=0.74$ $SD_{pers}=1.3$ $MW_{soc}=0.83$ $SD_{soc}=1.2$ $MW_{pub}=0.75$ $SD_{pub}=1.33$). Die Scores sind in Abbildung 7-38 dargestellt, die dazugehörigen p-Werte in Tabelle 7-1.

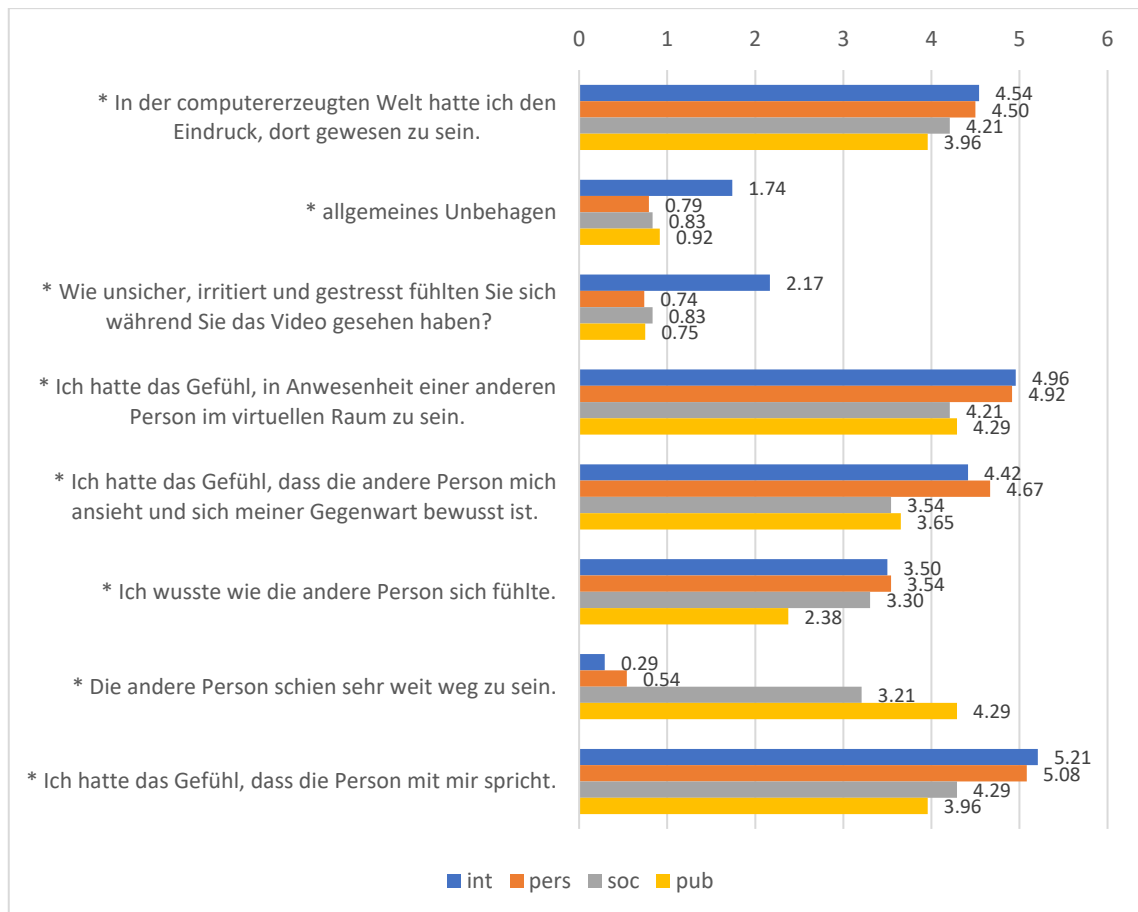


Abbildung 7-38: Scores für Presence und Wohlbefinden mit signifikanten Unterschieden zwischen den Testbedingungen (Likert-Skala von 0 bis 6).

	dort zu sein	allgemeines Unbehagen	irritiert gestresst	mich ansieht	sich fühlte	weit weg	mit mir spricht
int/pers	0.91	0.04	0.01	0.63	0.94	0.30	0.77
int/soc	0.45	0.04	0.01	0.16	0.80	0.00	0.04
int/pub	0.15	0.07	0.02	0.21	0.03	0.00	0.01
pers/soc	0.33	0.92	0.75	0.03	0.37	0.00	0.04
pers/pub	0.04	0.76	1.00	0.03	0.04	0.00	0.01
soc/pub	0.45	0.83	0.84	0.75	0.03	0.02	0.38

Tabelle 7-1: p-Werte für die in Abbildung 7-38 dargestellten Items. Signifikant unterschiedliche Werte sind fett dargestellt.

Wesentlich mehr Teilnehmer empfanden ein allgemeines Unbehagen, waren irritiert und gestresst, wenn die Sprecherin sehr nah war (int-Fall). Sowohl im int- wie auch im pers-Fall hatten die Teilnehmer das Gefühl, dass die Person im Film mit ihnen spricht. Die beiden Fälle unterschieden sich dabei untereinander nicht signifikant (pint/pers=0.78), allerdings zu den beiden anderen Fällen (p-Werte zwischen 0.01 und 0.04)

Experience

Mit der *int*-Distanz ($MW_{int}=3.17$ $SD_{int}=2.06$) wurde die Sprecherin als aufdringlicher empfunden als bei den anderen Distanzen (p -Werte < 0.001). Sowohl bei der *int*-Distanz ($MW_{int}=2.08$ $SD_{int}=2.06$) als auch bei der *pub*-Distanz ($MW_{pub}=2.17$ $SD_{pub}=1.97$) fühlten sich die Teilnehmer abgelenkter als unter den anderen beiden Bedingungen (Abbildung 7-39).

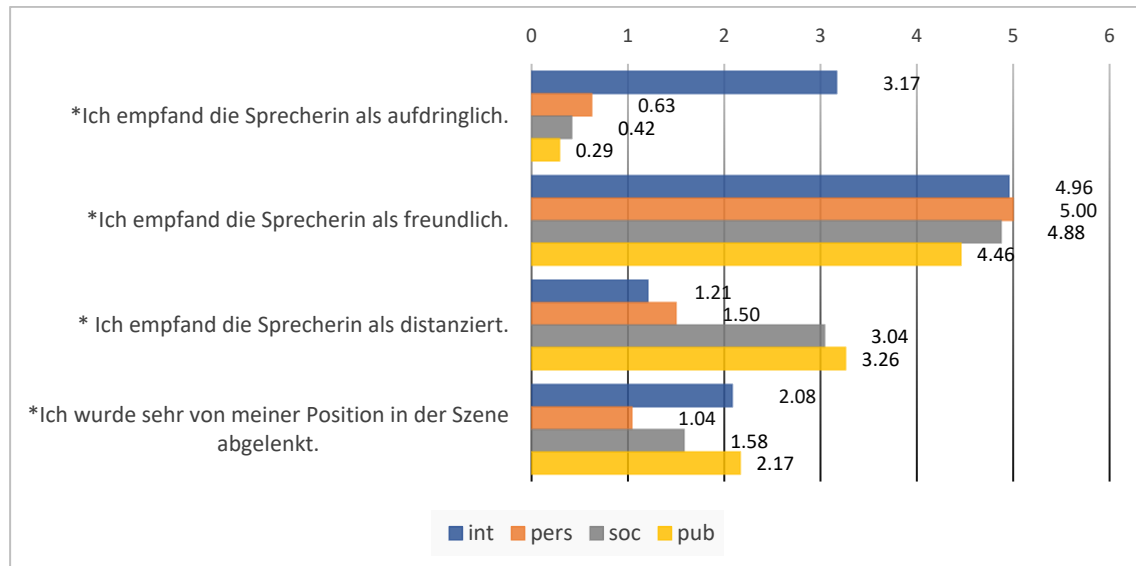


Abbildung 7-39: Scores der Items mit signifikant unterschiedlichen Ergebnissen (Likert-Skala von 0 bis 6).

User Experience

Die Teilnehmer empfanden das Video der *pers*-Bedingung am angenehmsten ($MW_{pers}=3.54$ $SD_{pers}=1.26$) und natürlichsten ($MW_{pers}=4.38$ $SD_{pers}=1.01$). Allerdings wurde das Video in der *soc*-Distanz als spannender empfunden ($MW_{soc}=3.88$ $SD_{soc}=1.39$). Die *pub*-Distanz schnitt in allen Aspekten am schlechtesten ab, für das Item *nervig/angenehm* gemeinsam mit der *int*-Distanz. Die Ergebnisse sind in Abbildung 7-40 dargestellt, die dazugehörigen p -Werte in Tabelle 7-2.

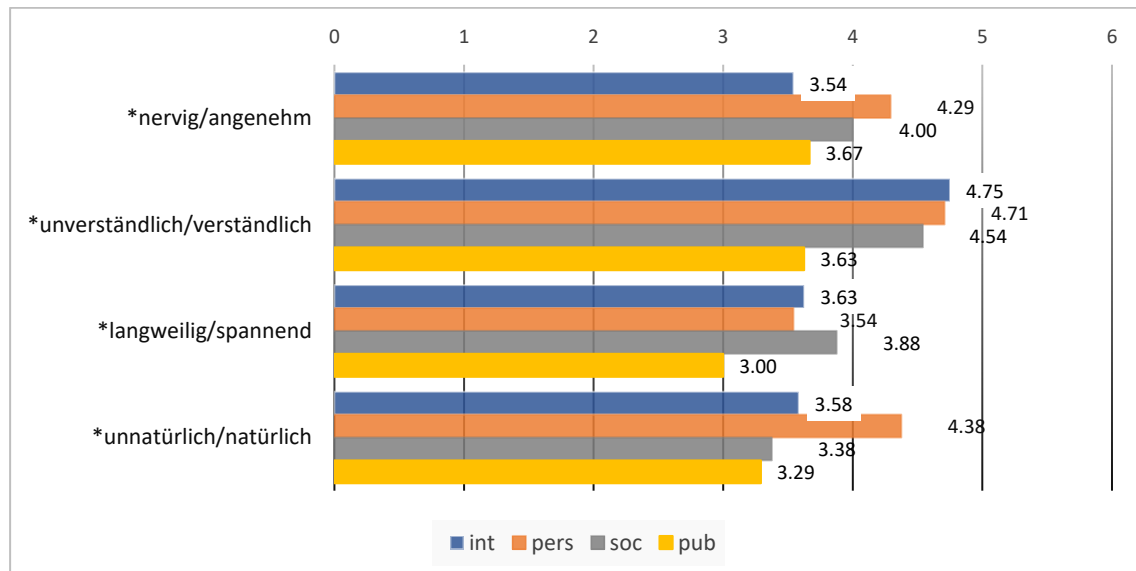


Abbildung 7-40: Scores der Antworten auf die Frage „Wie hast du das Video empfunden?“ (Likert-Skala von 0 bis 6).

	nervig /angenehm	unverständlich /verständlich	langweilig /spannend	unnatürlich /natürlich
int/pers	0.08	0.9044	0.81	0.027
int/soc	0.18	0.5526	0.47	0.58
int/pub	0.67	0.0041	0.1	0.443
pers/soc	0.34	0.6274	0.39	0.002
pers/pub	0.04	2E-05	0.11	0.002
soc/pub	0.16	0.0228	0.01	0.775

Tabelle 7-2: p-Werte für die in Abbildung 7-40 dargestellten Items. Signifikant unterschiedliche Werte sind fett dargestellt.

7.2.1.4 Diskussion

Die Ergebnisse zeigen, dass unterschiedliche Kameraabstände verschiedenartige Empfindungen bei den Zuschauern hervorrufen. In unserer Szenerie, in der jemand Informationen zu den Exponaten in einem Museum gibt, wurde eine nahe Distanz von 0.45 m als aufdringlich empfunden und eine weite Distanz von 5 m als weniger spannend. Am natürlichsten und angenehmsten empfanden die Teilnehmer die persönliche Distanz (1 m). In den beiden nahen Distanzen (*int*, *pers*) hatten die Teilnehmer ein intensiveres Gefühl, dass die Person im Video sie ansieht und zu ihnen spricht, als in den beiden entfernten Distanzen (*soc*, *pub*).

Sollen die Zuschauer, das Gefühl haben, dass die Person vor der Kamera zu ihnen spricht, scheint eine persönliche Distanz (in unserer Studie 1 m) gut geeignet zu sein, ohne beim Zuschauer negative Empfindungen hervorzurufen. Eine intime Distanz wird auch in CVR als aufdringlich empfunden und erzeugt Unbehagen. In anderen Szenarien, als der von uns verwendeten, kann dies aber durchaus gewollt sein und als Story-Element eingesetzt werden. Sowohl sehr nahe (*int*), als auch sehr weite Distanzen (*pub*) wirken auf den Zuschauer unnatürlicher und lenken den Zuschauer eher ab als die anderen beiden getesteten Abstände.

7.2.2 Studie: F-Formationen

Die Studie ist Teil der Bachelorarbeit von Vanessa Sarakiotis (2018): Space Composition in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit der Studentin die Idee entwickelt, die sie implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir von der Studentin durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Auch wenn in CVR der Filmmacher nicht den genauen Bildausschnitt bestimmen kann, kann er diesen durch die Komposition des Raumes beeinflussen. Für die Aufnahme eines Dialoges gilt es zum Beispiel zu entscheiden, ob Personen so nah beieinander sind, dass sie gleichzeitig im FoV wahrgenommen werden können, oder so weit voneinander entfernt sind, dass der Betrachter den Kopf bewegen muss, um beide zu sehen. In unserer Studie haben wir drei Anordnungen verglichen.

close-Seitenansicht

Die Charaktere sind nahe beieinander, und der Betrachter kann beide gemeinsam sehen, wenn der richtige Bildausschnitt ausgewählt ist (Abbildung 7-41, links; Abbildung 7-42). Der Abstand zwischen den Charakteren beträgt ca. 1 m (persönlicher Raum). Die Kamera ist 2 m von ihnen entfernt. Die Charaktere sind von der Seite sichtbar.

wide-Seitenansicht

Die Charaktere sind weiter voneinander entfernt und nur getrennt zu sehen (Abbildung 7-41, Mitte; Abbildung 7-43). Die Figuren sind im sozialen Raum des Betrachters in einer virtuellen Entfernung von ca. 2 m platziert. Auch hier sind die Charaktere von der Seite zu sehen.

Frontalansicht

Der Betrachter befindet sich in einem Abstand von jeweils 1 m zwischen den Charakteren (Abbildung 7-41, rechts; Abbildung 7-44). Eine Rotation von 180° ist nötig, um den Blick zwischen ihnen zu wechseln. Die Figuren sind von vorn sichtbar.

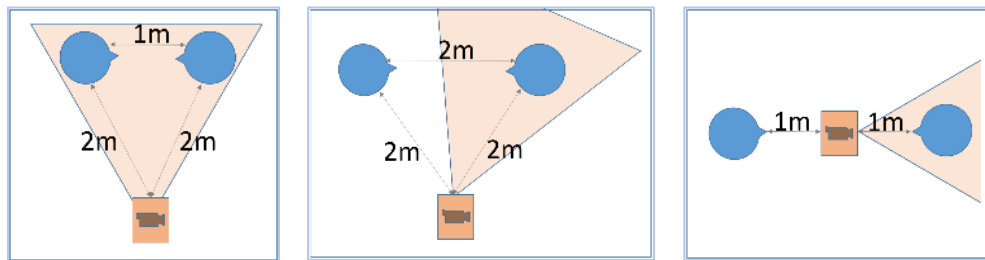


Abbildung 7-41: Die drei Ansichten: *close-Seitenansicht* (links), *wide-Seitenansicht* (Mitte), *Frontalansicht* (rechts).

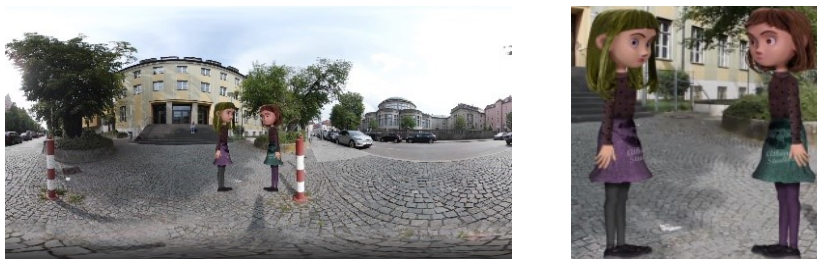


Abbildung 7-42: *close-Seitenansicht* - Die beiden Charaktere sind nahe beieinander, und der Betrachter kann sie gleichzeitig sehen. Sie sind von der Seite sichtbar. Links: die Projektion des Films. Rechts: das ungefähre FoV.



Abbildung 7-43: *wide-Seitenansicht* - Die Charaktere sind weiter voneinander entfernt. Der Betrachter muss den Kopf bewegen, um zwischen ihnen zu wechseln. Die Charaktere sind von der Seite sichtbar. Links: die Projektion des Films. Rechts: das ungefähre FoV.



Abbildung 7-44: *Frontalansicht* - Der Betrachter befindet sich zwischen den beiden Charakteren und muss den Kopf drehen, um zwischen ihnen zu wechseln. Die Figuren sind von vorne zu sehen. Links: Projektion des Films. Rechts: das ungefähre FoV.

7.2.2.1 Material

Für die Videos wurden animierte Charaktere verwendet. Dies bot die Möglichkeit, nahezu identisches Videomaterial für die verschiedenen Ansichten zu produzieren und die Position der Figuren leicht zu variieren. Mit menschlichen Schauspielern müsste derselbe Film mehrmals aufgenommen werden. Dadurch können kleine inhaltliche Abweichungen entstehen, die möglicherweise das Ergebnis beeinflussen. Außerdem wären Positionsänderungen nach Drehschluss schwer zu realisieren. Es wurden drei Videos produziert, jedes mit der gleichen Geschichte, bestehend aus nur einer Einstellung. Der einzige Unterschied zwischen den Videos war die Position der Charaktere.

Es wurden bereits existierende Charaktere von Athena Studios (Athena Studios, 2019) verwendet und diese mit Adobe Character Animator CC animiert. Die Stimmen wurden separat in Stereo ohne Hintergrundgeräusche aufgenommen. Dies ermöglichte es, den Sound räumlich zu produzieren, so dass die Stimmen immer aus der Richtung des sprechenden Charakters kamen. Zur Realisierung des räumlichen Sounds wurde REAPER v5.90 verwendet. Die Synchronisation von Sprache und Lippen der Charaktere wurde von Adobe Character Animator CC berechnet und manuell korrigiert. Mittels dieser Software wurden auch die Bewegungen der Figuren animiert: Die Desktop-Kamera zeichnete die Mimik und Bewegungen des Software-Benutzers auf und passte die animierte Figur an.

Für die Hintergrundaufnahme wurde eine Samsung Gear 360 Kamera genutzt (Auflösung 3.480 x 1.920, 30 fps). Die beiden animierten Charaktere, der räumliche Sound und der Hintergrund wurden in Adobe Premiere Pro CC importiert, zusammengefügt und die Filme als MP4-Dateien erstellt. Die Geschichte hatte eine Länge von 4 Minuten und wurde in 3 Teile von je 1:20 Minuten aufgeteilt, so dass jeder Teilnehmer alle drei Anordnungen in der Studie sehen konnte. Alle Teilnehmer sahen sich die Videos mit einer Samsung Gear VR an, wobei die Richtung des Kopfes aufgezeichnet wurde.

7.2.2.2 Pilotstudie

Um die besten Parameter für die Studie zu finden, führten wir eine Pilotstudie mit 6 Teilnehmern (1 weiblich, 5 männlich) durch. Alle von ihnen hatten Erfahrungen mit HMDs und CVR. Es wurden Parameter, wie die Größe der Hauptfiguren, ihr Abstand voneinander und ihre Position in der omnidirektionalen Szene getestet und angepasst.

7.2.2.3 Teilnehmer und Prozedur

An der Hauptstudie nahmen 24 Personen (8 weiblich, 16 männlich, Alter: MW=27.7 SD=6.9) teil. 75 % von ihnen hatten keine oder nur wenig CVR-Erfahrung.

Es wurde ein within-subject Testdesign ausgewählt, bei dem jeder Teilnehmer alle drei Anordnungen sah. Die unabhängige Variable war die Anordnung von Kamera und Charakteren (*close-Seitenansicht*, *wide-Seitenansicht*, *Frontalansicht*). Die Teilnehmer sahen sich drei Videoteile mit jeweils einer anderen Anordnung an, wobei die Reihenfolge der Anordnungen permutiert wurde.

Nach jedem Filmabschnitt wurden per Fragebogen ermittelt:

- Task Workload (NASA TLX) (Hart und Staveland, 1988)
- Aufmerksamkeit (7-Punkte-Likert-Skala)
- Enjoyment (7-Punkte-Likert-Skala)
- User Experience (UEQ, short) (Hinderks u. a., 2018)
- Presence (IPQ) (Schubert u. a., 2002)
- Simulator-Sickness (SSQ) (Kennedy u. a., 1993)

Der Fragebogen endete mit Fragen zum direkten Vergleich der Anordnungen, wobei mehr als nur eine Antwort gewählt werden konnte. Die Teilnehmer konnten ihren favorisierten Clip auswählen und sollten Gründe für ihre Wahl angeben. Die Studie dauerte ca. 45 Minuten pro Teilnehmer.

7.2.2.4 Ergebnisse

Der Shapiro-Wilk-Test zeigte, dass nicht für jedes Item Normalverteilung vorausgesetzt werden kann. Deshalb wurde der Friedman-Test verwendet, um signifikante Unterschiede zwischen den Anordnungen zu finden. Zur Bestimmung der Signifikanz wurde ein α -Level von 0.05 gewählt.

Presence und Simulator-Sickness

Es gab keine signifikanten Unterschiede in Bezug auf Presence und Simulator-Sickness.

Task Workload

Für zwei Items des NASA-TLX unterschieden sich die Werte signifikant. In beiden Fällen war die Belastung für die *Frontalansicht* höher als für die *close*-Seitenansicht: körperliche Belastung ($MW_{close}=20.44$ $SD_{close}=14.29$ $MW_{front}=33.91$ $SD_{front}=34.24$ $p = 0.03$) und allgemeine Belastung ($MW_{close}=16.73$ $SD_{close}=18.06$ $MW_{front}=29.35$ $SD_{front}=30.01$ $p = 0.04$).

Aufmerksamkeit

Zur Überprüfung der Aufmerksamkeit wurden zehn Fragen zum Filminhalt gestellt. Die meisten korrekten Antworten wurden für die *wide*-Seitenansicht (81.25 %) gegeben, gefolgt von der *close*-Seitenansicht (78.7 %). Die *Frontalansicht* erbrachte die wenigsten richtigen Antworten (71.25 %).

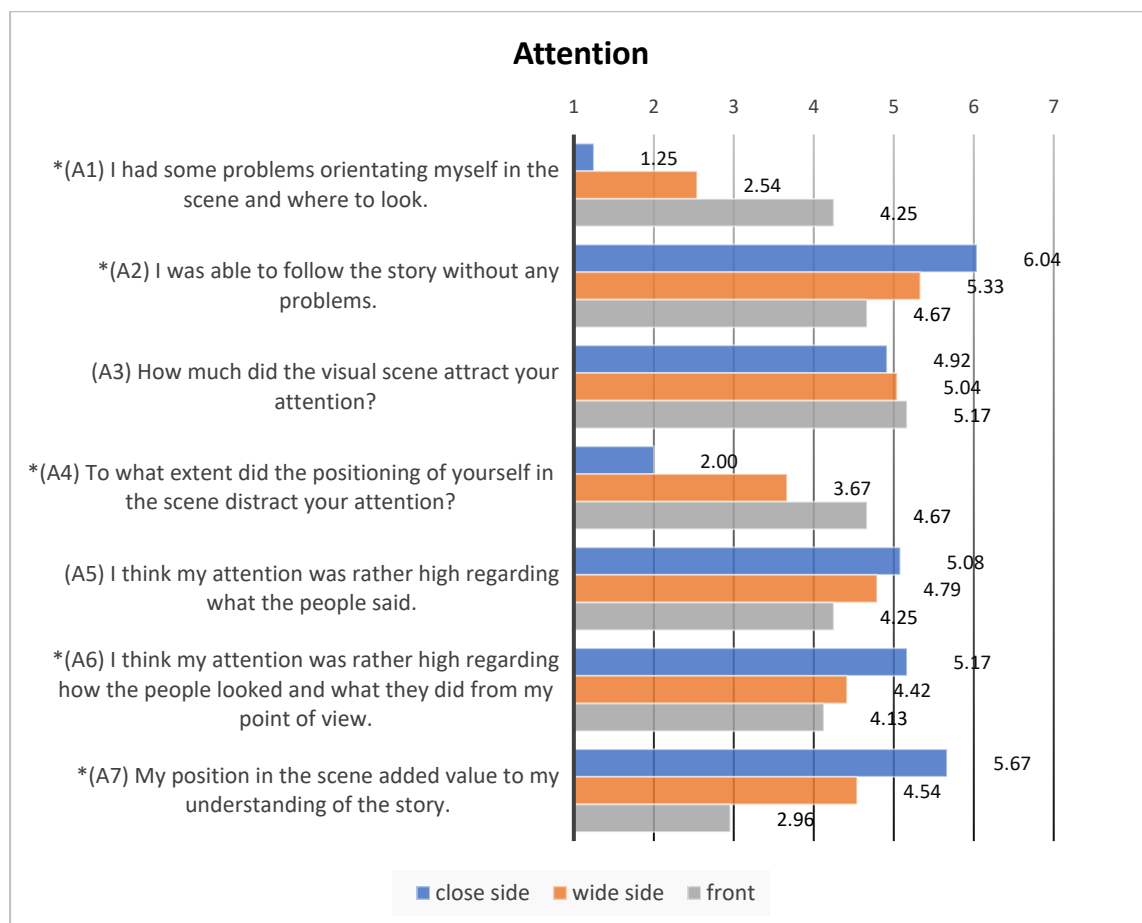


Abbildung 7-45: Mittelwerte für die selbsteingeschätzte Aufmerksamkeit (Likert-Skala von 1 bis 7). Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

Bei den selbst eingeschätzten Bewertungen der Aufmerksamkeit gab es bei fünf der sieben Fragen Unterschiede zwischen den Ansichten. Bei all diesen Fragen ergab die *close*-Seitenansicht die beste und die *Frontalansicht* die niedrigste Punktzahl. Abbildung 7-45 zeigt die Scores und Tabelle 7-3 die p-Werte der signifikanten Ergebnisse der paarweisen post-hoc Tests.

	A1	A2	A4	A6	A7
close-wide	0.008	0.18	<0.001	0.13	0.025
close-front	<0.001	0.017	<0.001	0.03	<0.001
wide-front	0.001	0.19	0.03	0.6	0.01

Tabelle 7-3: Die p-Werte der paarweisen post-hoc Tests für Fragen mit signifikanten Unterschieden. Fettgedruckte Zahlen bedeuten signifikante p-Werte.

Enjoyment

Bei vier der sechs Enjoyment-Fragen fanden wir signifikante Unterschiede in den Daten. Auch hier wurde die *close*-Seitenansicht am besten bewertet, und die *Frontalansicht* führte zu den niedrigsten Scores. Abbildung 7-46 zeigt die Ergebnisse und Tabelle 7-4 die p-Werte für die Fragen mit signifikanten Unterschieden zwischen den Anordnungen.

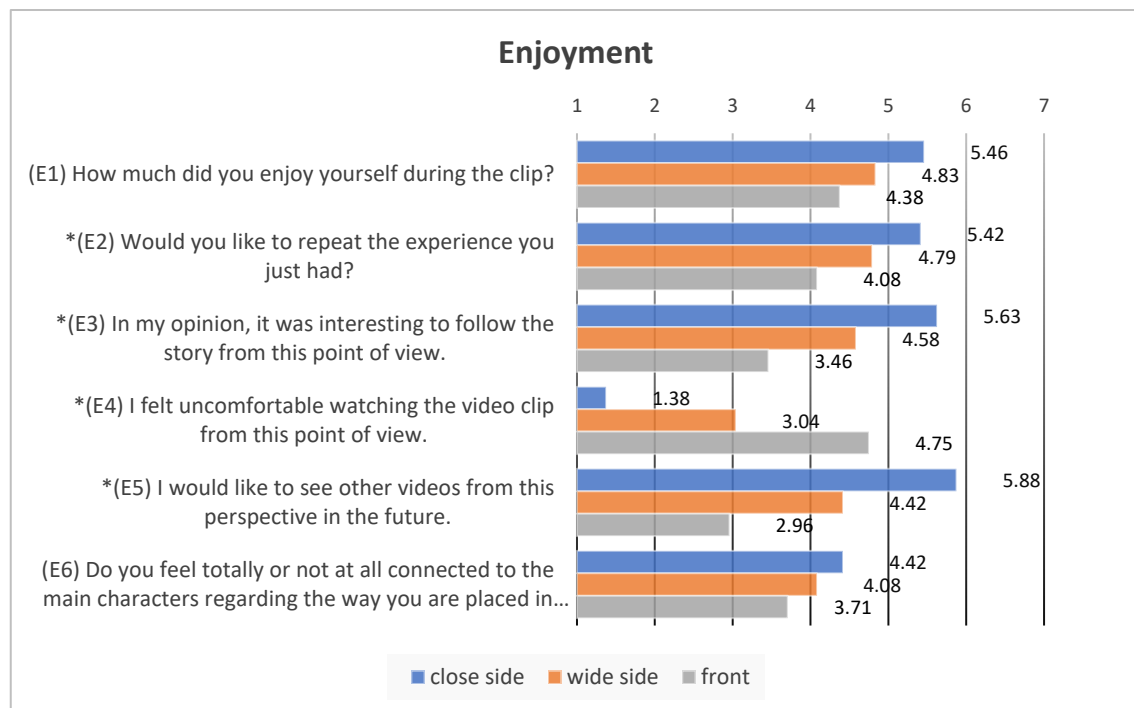


Abbildung 7-46: Mittelwerte für die Enjoyment-Items (Likert-Skala von 1 bis 7). Die close-Seitenansicht erzielte die besten Ergebnisse. Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

	E2	E3	E4	E5
close-wide	0.19	0.09	0.002	0.01
close-front	0.025	0.005	<0.001	<0.001
wide-front	0.14	0.03	0.004	0.02

Tabelle 7-4: Die p-Werte der paarweisen post-hoc Tests für Fragen mit signifikanten Unterschieden. Fettgedruckte Zahlen bedeuten signifikante p-Werte.

User Experience

Auch für das Benutzererlebnis wurde die *close*-Seitenansicht im Vergleich zu den beiden anderen Anordnungen am besten bewertet. Abbildung 7-47 zeigt die Ergebnisse des UEQs und Tabelle 7-5 die p-Werte für die Items, bei denen signifikante Unterschiede zwischen den Anordnungen festgestellt wurden.

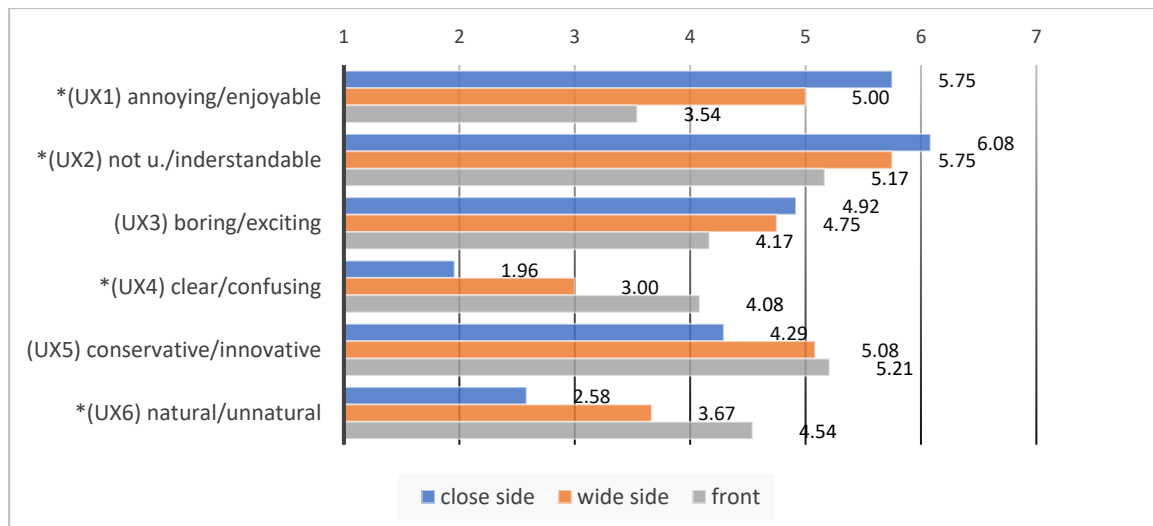


Abbildung 7-47: Mittelwerte für die UX-Items. Die close-Seitenansicht führte zu besseren Ergebnissen (Likert-Skala von 1 bis 7). Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

	UX1	UX2	UX4	UX6
close-wide	0.1	0.45	0.05	0.016
close-front	0.001	0.04	<0.001	0.001
wide-front	0.015	0.21	0.05	0.06

Tabelle 7-5: Die p-Werte der paarweisen post-hoc Tests für Fragen mit signifikanten Unterschieden. Fettgedruckte Zahlen bedeuten signifikante p-Werte.

Vergleich

Der Vergleich der drei Ansichtsarten nach dem Ansehen aller Videos wies die besten Ergebnisse für die *close*-Seitenansicht auf. Die meisten Teilnehmer (45.8 %) fühlten sich bei dieser Ansicht am präsentesten und waren am aufmerksamsten (50 %). Gründe dafür sind: „weil ich beide Charaktere gleichzeitig gut gesehen habe“ (P16), „es war einfacher, dem Gespräch zu folgen“ (P18), „man braucht nicht die ganze Zeit den Kopf zu drehen“ (P16) und es war „die natürlichste Sicht“ (P4). Die quantitativen Ergebnisse dieser Fragen sind in Abbildung 7-48 dargestellt.

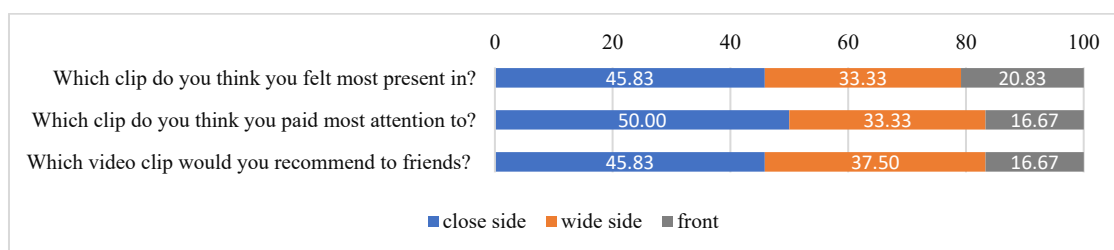


Abbildung 7-48 Anteil der Teilnehmer, die für die jeweilige Anordnung gestimmt haben (in Prozent). Die *close*-Seitenansicht wurde in allen drei abschließenden Fragen bevorzugt.

Auch wenn die Frontansicht zu relativ niedrigen Scores führte, ist es überraschend, dass einige Teilnehmer diese Ansicht bevorzugten, da es sehr ungewöhnlich ist, zwischen zwei sprechenden Personen platziert zu werden. Diese Erfahrung entspricht nicht der Realität und ist mit mehr Aufwand für das Bewegen des Kopfes verbunden. Gründe für die positive Bewertung sind „Ich fühlte mich aufgrund der Position in die Geschichte involviert“ (P6), „weil ich dachte, das Mädchen redet mit mir“ (P24), „es ist auf diese Weise anstrengender, aber die magischen Kräfte haben den größten Effekt“ (P21). Die Annahme, dass es eine Korrelation zwischen diesem Ergebnis und der VR-Erfahrung des Zuschauers gibt, wurde durch die Daten nicht bestätigt.

Headtracking-Daten

Mit dem CVR-Analyzer Tool (Rothe et al., 2018a) wurden die Headtracking-Daten visualisiert und Heatmaps generiert. Wie zu erwarten, waren die Kopfrichtungen in der *close-Seitenansicht* (Abbildung 7-49) konzentrierter und in den beiden anderen Ansichten verteilter (Abbildung 7-50 und Abbildung 7-51).

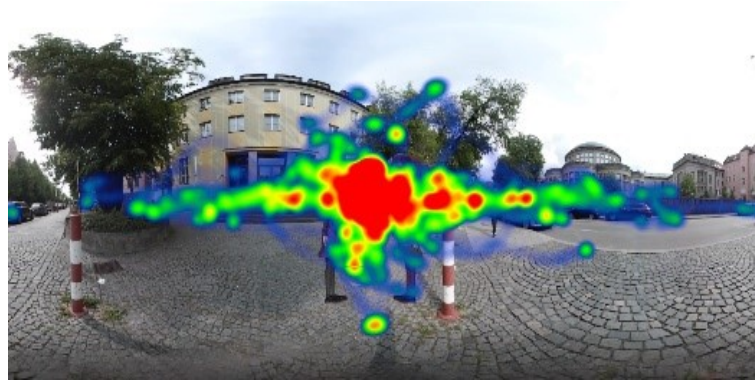


Abbildung 7-49: *close-Seitenansicht* - Die Teilnehmer können beide Charaktere gleichzeitig sehen, wenn die Kopfrichtung zwischen den Charakteren zentriert ist. Die Heatmap zeigt, dass die Teilnehmer in diese Richtung fokussiert waren.

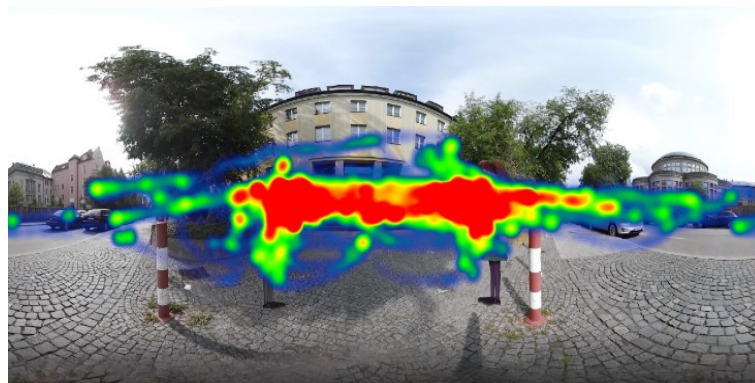


Abbildung 7-50: *wide-Seitenansicht* - Der Betrachter muss die Kopfrichtung ändern, um die Charaktere zu sehen. Die Heatmap zeigt, dass die Teilnehmer insbesondere die Bereiche der Charaktere und zwischen ihnen inspizieren.

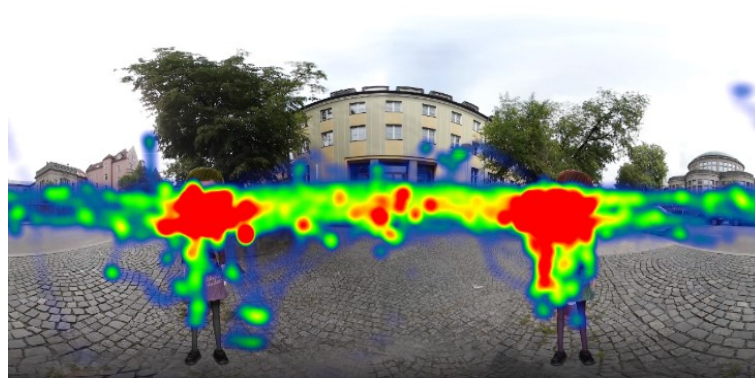


Abbildung 7-51: *Frontalansicht* - Der Betrachter befindet sich in der Mitte zwischen den beiden Charakteren und muss den Kopf um 180° drehen, um zwischen ihnen zu wechseln. Die Figuren sind von vorne zu sehen. Die Heatmap zeigt, dass die Teilnehmer die Bereiche der Charaktere inspizieren, aber auch den Bereich dazwischen.

Bei der genaueren Untersuchung der Tracking-Daten für kleinere Intervalle, wurde beobachtet, wie die Teilnehmer von einem Charakter zum anderen wechselten. Dabei fiel auf, dass die Betrachter bei beiden Seitenmodi das Gesicht der Charaktere nicht in der FoV zentrierten (Abbildung 7-52, links). Stattdessen entspricht der ausgewählte Bildausschnitt den Regeln des Goldenen Schnitts (Abbildung 7-52, rechts). Dieses Ergebnis kann auch dadurch beeinflusst sein, dass mehr Kopfbewegungen nötig wären, um genau das Gesicht zu fixieren.



Abbildung 7-52: *wide-Seitenansicht* - die rechte Person spricht. Die Heatmap zeigt, dass die Teilnehmer das Gesicht nicht im FoV zentriert haben. Das gewählte Bild folgt den Regeln des Goldenen Schnitts (rechts).

7.2.2.5 Diskussion und Einschränkungen

In unserer Nutzerstudie wurde die *close-Seitenansicht* von den meisten Teilnehmern bevorzugt. Dies kann dadurch verursacht sein, dass das Ändern der Blickrichtung zwischen den Charakteren weniger Aufwand erfordert. Diese Ansicht ist bequemer als die anderen beiden Anordnungen, und die Charaktere können gleichzeitig gesehen werden.

Allerdings wurde auch die ungewöhnliche Position zwischen den beiden Charakteren von einigen Teilnehmern akzeptiert. Sie berichteten von einem höheren Presence-Gefühl, was auch durch den geringeren Abstand zwischen Kamera und Charakter verursacht sein kann.

Mit den aufgezeichneten Headtracking-Daten konnten wir die Kopfbewegungen untersuchen. Die Teilnehmer konzentrierten sich zumeist auf die sprechende Person, zentrierten diese jedoch im Allgemeinen nicht im FoV. Stattdessen wurde die Kopfrichtung seitlich des Charakters gewählt, was zu Bildern führte, die den Regeln des Goldenen Schnitts folgen. Dies muss bei der Komposition von Szenen berücksichtigt werden.

Alle unsere Ergebnisse basieren auf Kurzfilmen mit zwei Figuren und dem Betrachter / der Kamera in drei verschiedenen F-Formationen (Abbildung 7-41). Es bedarf weiterer Untersuchungen, um allgemeine Regeln für verschiedene Szenarien zu finden. Wie der Betrachter in solchen Szenarien reagiert, hängt nicht nur von der Kameraentfernung, der Formation der Figuren und der Kamera ab. Auch Aspekte wie Umgebung oder Hintergrundgeräusche sind relevant. In unserem Film war der visuelle Hintergrund für die Geschichte diskret und unwichtig. Die Stimmen der Charaktere waren räumlich und hoben sich deutlich vom nicht-räumlichen Umgebungsgeschall ab.

Während des Gesprächs änderte sich die Sprecherrolle nicht sehr oft. Die Charaktere sprachen zwischen 5 und 17 s, bevor das Gespräch zur anderen Person wechselte. Der Betrachter hatte genügend Zeit, um die Richtung zu ändern. Denkbar sind aber auch schnellere Gespräche, bei denen schon nach wenigen Sekunden der Sprecher wechselt und der Betrachter nicht so schnell die Blickrichtung ändern kann. Zusätzliche Untersuchungen sind notwendig um herauszufinden, wie sich Zuschauer in solchen Situationen verhalten. Andererseits ist es in einem langsameren Gespräch, in dem eine Person länger spricht, vorstellbar, dass der Betrachter die Umgebung inspiziert, nachdem er das Aussehen der Figur bereits kennt.

Zusätzlich werden die Ergebnisse unserer Studie von der Größe des FoV beeinflusst, die von verschiedenen Parametern abhängt, z. B. dem HMD und der persönlichen Augenpositionen.

7.2.2.6 Fazit

In dieser Studie wurde das Betrachtungsverhalten für ein Szenario (zwei stehende und sprechende Personen) analysiert. Weitere Untersuchungen anderer Szenarien mit mehr Personen in anderen Konstellationen sind erforderlich. Wie fühlt sich der Betrachter in einer Gruppe, als Teil der Gruppe oder als unsichtbarer Beobachter? Unter welchen Umständen beeinflusst die Distanz solche Gefühle? Darüber hinaus wird mehr Kenntnis über die Auswirkungen von Hintergrund, Sprechdauer und Sound benötigt.

7.2.3 Studie: Diskrepanzen in der Körpergröße

Die Studie ist Teil der Bachelorarbeit von Boris Kegeles (2017): Influences of Camera Position to Attention and Presence in CVR.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Die Ergebnisse dieser Studie wurden vorab veröffentlicht in:

- Rothe, S., Kegeles, B., Hussmann, H., 2019. Camera Heights in Cinematic Virtual Reality, in: Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '19. ACM Press, New York, New York, USA
- Rothe, S., Kegeles, B., Hussmann, H., 2019. Camera Heights in Cinematic Virtual Reality, in: Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '19. ACM Press, New York, New York, USA

Ziel dieser Studie war es herauszufinden, wie unterschiedliche Kamerahöhen vom Zuschauer wahrgenommen werden.

Die Studie ist in drei Teile gegliedert:

1. In einer VR-Umgebung wurde die eigene Körpergröße geschätzt.
2. Sitzend schauten die Teilnehmer Videos an, die in verschiedenen Sitz-Höhen aufgenommen wurden.
3. Stehend schauten die Teilnehmer Videos an, die in verschiedenen Steh-Höhen aufgenommen wurden.

7.2.3.1 Material

Um herauszufinden, wie die Benutzer in einer VR-Umgebung ihre Körpergröße empfinden, wurde eine Anwendung mit Unity 3D und dem Pack Gesta Furniture von Gesta2 implementiert. Es wurde ein VR-Wohnzimmer mit Möbelstücken in standardisierten Größen als Höhenreferenzobjekte kreiert (Abbildung 7-53). Daraus wurde für die Verwendung der Samsung Gear VR eine Android-App generiert. Mit einem Bluetooth-Controller konnten die Teilnehmer ihre Höhe im VR-Raum so lange ändern, bis sich die Blickhöhe in VR wie ihre eigene Augenhöhe anfühlte. Die Teilnehmer konnten sich nicht im virtuellen Raum bewegen und befanden sich alle am gleichen Ort.



Abbildung 7-53: VR-Wohnzimmer zur Identifizierung der eigenen gefühlten Körpergröße in einer VR-Umgebung (erstellt mit Unity 3D und dem Gesta Furniture Pack).

Um zu untersuchen, wie die Höhe der Kamera Presence, Simulator-Sickness und Experience beeinflusst, wurden kurze Videos mit unterschiedlichen Höhen der Kamera produziert. Dazu wurden zwei typische Situationen - eine im Stehen, die andere im Sitzen – ausgewählt: Das erste Video wurde in einer sitzenden Position an einem Tisch in einem Café aufgenommen (Abbildung 7-54, links). Ein Ehepaar führt ein Gespräch mit einem jungen Mann; der Zuschauerplatz ist ebenfalls an diesem Tisch. Diese Videos nennen wir *Sitzvideo*. Die Szene wurde für verschiedene Höhen der Kamera erstellt (100 cm, 110 cm, 120 cm, 130 cm). Die Situation ist in allen Videos gleich, nur der Text der Konversation ist ein anderer. Das zweite Video wurde in einer stehenden Position in einer Fußgängerzone aufgenommen (Abbildung 7-54, rechts). Ein Fußballkünstler jongliert mit einem Ball, während Fußgänger zuschauen. Wir nennen dieses Video *Stehvideo*. Die Szene wurde ebenfalls für verschiedene Höhen der Kamera produziert: 150 cm, 160 cm, 170 cm, 180 cm, 190 cm und 200 cm. Die Situation ist in allen Videos gleich. Es gibt jedoch unterschiedliche Performances des Fußballspielers und das Publikum ändert sich.



Abbildung 7-54: Das verwendete Videomaterial, links: Video in einem Café, aufgenommen in sitzender Höhe (die Gesichtsschwärzung war während der Studie nicht vorhanden), rechts: Video in einer Fußgängerzone, aufgenommen in stehender Höhe.

7.2.3.2 Teilnehmer und Prozedur

26 Personen (9 weiblich, 17 männlich, Alter $MW=33.2$, $SD = 15.4$) nahmen an der Studie teil. 73.1 % der Teilnehmer hatten VR-Erfahrungen. Vor Beginn der Tests wurde die Körpergröße und die Augenhöhe der Teilnehmer im Stehen und Sitzen gemessen. Die durchschnittliche Körpergröße der Teilnehmer betrug 175.4 cm ($SD = 8.89$), die durchschnittliche Augenhöhe in stehender Haltung 165.3 cm ($SD = 8.91$) und in sitzender Haltung 122.1 cm ($SD = 4.7$). Der Stuhl hatte eine Höhe von 45 cm. Bevor die Teilnehmer den Film sahen, wurden einige demographische Fragen gestellt. Teilnehmer ohne VR-Erfahrung erhielten ein Zeitfenster von 5 bis 20 Minuten, um eine VR-Umgebung auszuprobieren, die die Studie nicht betraf. Ziel war es, den Novelty-Effekt zu reduzieren. Die Studie besteht aus drei Teilen:

- **1. Teil:** In einer stehenden Haltung änderten die Teilnehmer ihre eigene Betrachtungshöhe in einem VR-Raum, bis sie das Gefühl hatten, dass die Höhe ihrer eigenen Größe entsprach. Dafür verwendeten sie einen Bluetooth-Controller.
- **2. Teil:** Auf einem Stuhl sitzend (45 cm) wurden zwei Videos gezeigt, die mit unterschiedlichen Kamerahöhen aufgenommen wurden, eines in der richtigen Größe des Teilnehmers (+ -5 cm), das

andere 10 cm höher oder niedriger. Die Szene des Videos entsprach einer Sitzsituation an einem Tisch in einem Café (Abbildung 7-54, links).

- **3. Teil:** In einer stehenden Haltung wurden drei Videos gezeigt, die mit unterschiedlichen Kamerahöhen aufgenommen wurden, eines in der richtigen Größe des Teilnehmers (+ -5 cm), die anderen beiden 10 cm bzw. 20 cm höher oder 10 cm bzw. 20 cm niedriger. Die Szene des Videos entsprach einer stehenden Situation in einer Fußgängerzone (Abbildung 7-54, rechts).

Die Teilnehmer wurden nach dem Zufallsprinzip in zwei Gruppen aufgeteilt und es wurde ein between-subject Testdesign angewendet. Eine Gruppe schaute sich Videos an, deren Setups höher oder gleich ihrer eigenen Augenhöhe waren (Abbildung 7-55, links). Wir nennen dies die *above*-Gruppe. Die *below*-Gruppe schaute Videos, die niedriger oder gleich ihrer eigenen Größe waren (Abbildung 7-55, rechts). Die Teilnehmer wurden nicht über die Unterschiede der Videos informiert.

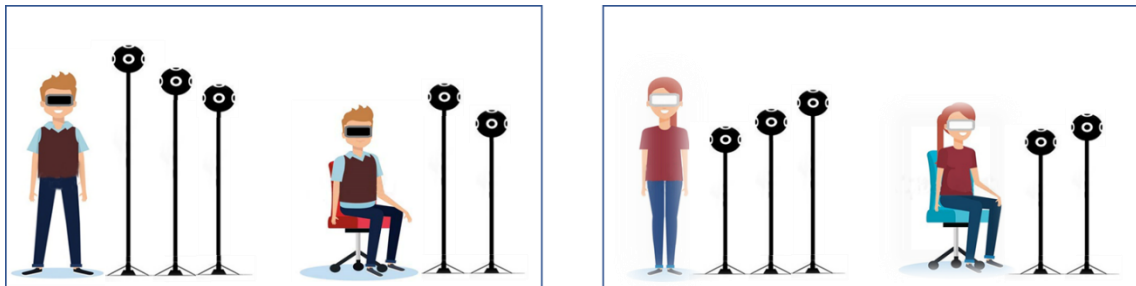


Abbildung 7-55: Testfälle für die beiden Gruppen, links: above-Gruppe, rechts: below-Gruppe.

Für die sitzende Variante schauten sich die Teilnehmer zunächst ein Video 10 cm über / unter ihrer Körpergröße und nach Beantwortung einiger Fragen ein Video für die richtige Sitzgröße (+ -5 cm) an.

Da die Augenhöhe von stehenden Personen stärker variiert als von sitzenden Personen, wurde in diesem Fall ein größeres Intervall untersucht. Für den stehenden Teil wurden Videos von 20 cm und 10 cm Unterschied zur Augenhöhe und zusätzlich ein Video in Augenhöhe angezeigt. Es wurde immer mit dem Unterschied von 20 cm begonnen, gefolgt von dem Unterschied von 10 cm und dem Video in Augenhöhe. Nach jedem Video wurden Teile des Fragebogens beantwortet, mit dem die folgenden abhängigen Variablen ermittelt wurden: Presence, Simulator-Sickness, Benutzererfahrung. Der Fragebogen endete mit Fragen zum Vergleich der beiden Methoden.

7.2.3.3 Ergebnisse

Erster Teil

Im ersten Teil der Studie bestimmten die Teilnehmer ihre eigene gefühlte Körpergröße in einem VR-Raum (Abbildung 7-56). 42.6 % der Teilnehmer kamen ihrer realen Größe relative nahe (+/- 10 cm). 80 % der Teilnehmer identifizierten die VR-Größe kleiner als ihre tatsächliche Größe, 50 % wählten ihre Größe zu klein (mehr als 10 cm zu klein). Nur zwei Personen (7.7 %) wählten diese zu groß (mehr als 10 cm zu groß). Beide Personen haben VR zum ersten Mal erlebt und waren sehr beeindruckt davon.

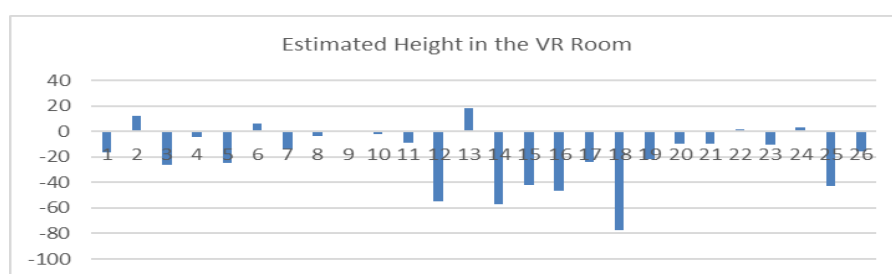


Abbildung 7-56: Das Bild zeigt die Unterschiede zwischen der geschätzten Höhe im VR-Raum und der tatsächlichen Höhe (in cm). Die meisten Teilnehmer bewerteten sich kleiner.

Zweiter Teil

Im zweiten Teil untersuchte die Studie den Einfluss unterschiedlicher Kamerahöhen in Sitzhaltungen.

Presence

Es wurden die folgenden 10 Fragen des Presence-Fragebogens (Schubert u. a., 2002) genutzt:

1. Ich hatte das Gefühl „da zu sein“. (allgemein)
2. Ich hatte das Gefühl, dass die virtuelle Welt mich umgab. (räumliche Präsenz)
3. Ich hatte das Gefühl, nur Bilder wahrzunehmen. (räumliche Präsenz)
4. Ich fühlte mich im virtuellen Raum nicht präsent. (räumliche Präsenz)
5. Wie bewusst waren Sie sich der realen Welt? (Involvement)
6. Ich war mir meiner realen Umgebung nicht bewusst. (Involvement)
7. Ich habe immer noch auf die reale Umgebung geachtet. (Involvement)
8. Wie real erschien dir das Video? (Realismus)
9. Inwieweit schien Ihre Erfahrung in der virtuellen Umgebung mit Ihrer realen Erfahrung übereinzustimmen? (Realismus)
10. Das Video schien realistischer als die reale Welt. (Realismus)

Die Presence war in den Fällen, in denen die Kamera zu niedrig war, höher als in denen mit zu hoher Kamera (Abbildung 7-57). Für zwei Items gab es einen signifikanten Unterschied: (3) $p = 0.083$ (6) ($p = 0.041$).

Beim Vergleich der Presence-Bewertungen für die Fälle, in denen die Kamera zu hoch war, und dem Fall, in der sich die Kamera auf Augenhöhe befand (+ -5 cm) (Baseline), ergaben sich für die Items (1), (4) und (5) signifikant höhere Werte bei der Baseline (7). Für die höhere Kamera war der Wert nur für Item (5) signifikant höher. Überraschenderweise gab es auch einige Unterschiede in der Baseline zwischen den beiden Gruppen. Für die Punkte (5) und (6) war die Presence-Bewertung für die *below*-Gruppe besser. Es bedarf weiterer Untersuchungen, ob dieses Ergebnis tatsächlich durch die Kamerahöhen in den vorhergehenden Videos verursacht wurde. Möglicherweise tritt eine Art interne Kalibrierung auf, wenn das erste Video angesehen wird.

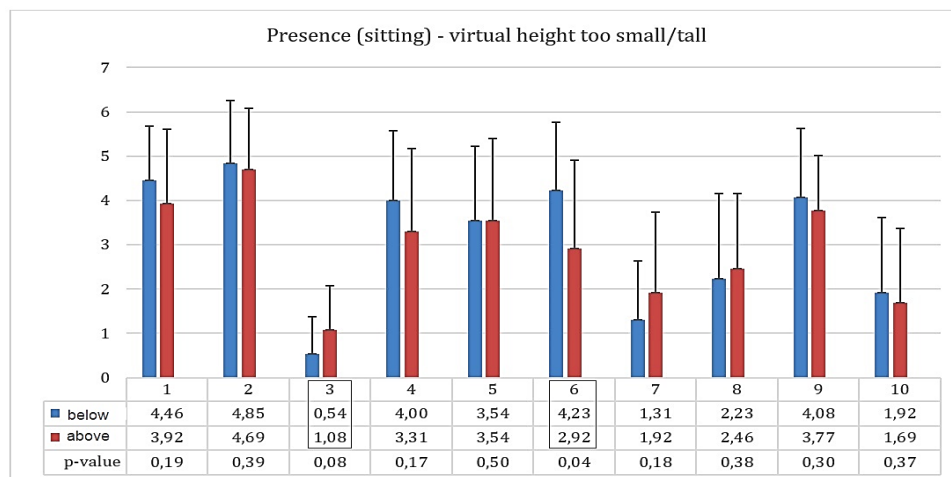


Abbildung 7-57: Das Bild zeigt die Mittelwerte und Standardabweichungen für die Presence-Fragen in einer Sitzhaltung, in der die Kameraposition 10 cm über oder unter der Augenhöhe liegt (Likert-Skala von 0 bis 7). Die Presence ist höher für den Fall, dass der Benutzer kleiner als in der realen Welt ist. Für die Items (3) und (6) bestehen signifikante Unterschiede.

Experience

Insgesamt gaben 53.8 % der Zuschauer seltsame Gefühle an, wenn die Kamera nicht ihrer Augenhöhe entsprach. 38.5 % in der *below*-Gruppe, 69.2 % in der *above*-Gruppe (Abbildung 7-58, links). Der exakte Fisher-Test ergab, dass dieser Unterschied signifikant ist ($p = 0.09$).

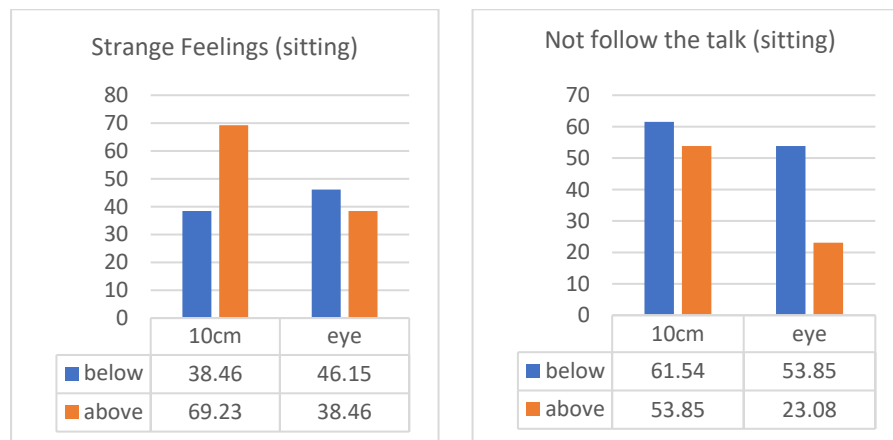


Abbildung 7-58: Die Bilder zeigen Unterschiede im Zuschauererlebnis (in Prozent). links: Ist die Kameraposition höher als die tatsächliche Augenhöhe, haben mehr Betrachter ein seltsames Gefühl als bei zu niedriger Kameraposition. Für die *below*-Gruppe gibt es keinen signifikanten Unterschied zur Baseline. rechts: Mehr als die Hälfte der Teilnehmer verpasste das Gespräch, wenn die Kamera zu hoch / niedrig war und sogar auf Augenhöhe, wenn die Kamera zuvor niedriger war. Mehr Teilnehmer verfolgten das Gespräch, wenn das vorherige Video zu hoch war.

Aufmerksamkeit

Vergleicht man die Aufmerksamkeit beim Ansehen des Videos, so gab es keinen signifikanten Unterschied zwischen den Gruppen, wenn die Kamera zu hoch / niedrig war (Abbildung 7-58, rechts). Auch hier gab es einen Unterschied zwischen den Gruppen in der Baseline-Bedingung. Die Teilnehmer verfolgten das Gespräch im Video auf Augenhöhe aufmerksamer, wenn sie zuvor ein höheres Video gesehen hatten.

Vergleich

69.2 % der Zuschauer sahen sich das Video lieber in ihrer eigenen Sitzhöhe an, 53.8 % in der *below*-Gruppe und 84.6 % in der *above*-Gruppe (Tabelle 7-6, oben). In der *above*-Gruppe erkannten alle Zuschauer richtig, welches Video sich in ihrer Sitzhöhe befand, in der anderen Gruppe 53.8 % (Tabelle 7-6, unten).

		+10 cm	eye level
Welches Video würden Sie bevorzugen?	<i>below</i>	46.2 %	53.8 %
	<i>above</i>	15.4 %	84.6 %
Welches Video entspricht Ihrer eigenen Augenhöhe?	<i>below</i>	46.2 %	53.8 %
	<i>above</i>	0 %	100 %

Tabelle 7-6: Oben: Mehr Zuschauer bevorzugten das Video in ihrer eigenen Augenhöhe, wenn die Kameraposition des vorherigen Videos zu hoch war. Unten: Mehr Zuschauer identifizierten das Video ihrer eigenen Augenhöhe, wenn das vorhergehende zu hoch war.

Für die *below*-Gruppe zeigten die Ergebnisse für beide Fragen keine signifikanten Unterschiede. Dies legt die Annahme nahe, dass der Betrachter in Sitzhaltungen es kaum wahrnimmt, wenn die Kamerahöhe 10 cm zu niedrig ist.

Dritter Teil

In diesem Teil sahen sich die Teilnehmer das Video stehend für drei Kamerahöhen an.

Presence

Es wurde ein Zwei-Stichproben t-Test durchgeführt und festgestellt, dass die allgemeine Presence für die *below*-Gruppe signifikant höher ist als für die *above*-Gruppe, wenn die Kamera von der Augenhöhe abweicht. Diese Aussage gilt sowohl für die 20 cm Differenz ($p = 0.026$) als auch für die 10 cm Differenz ($p = 0.044$) (Abbildung 7-59). Der Unterschied zwischen den Gruppen, wenn die Kamera der Augenhöhe entspricht, war nicht signifikant ($p = 0.106$). Des Weiteren ergaben sich in unseren Experimenten innerhalb der Gruppen keine signifikanten Unterschiede bezüglich der Presence zwischen der zu hohen Kamera und der Baseline sowie der zu niedrigen Kamera und der Baseline.

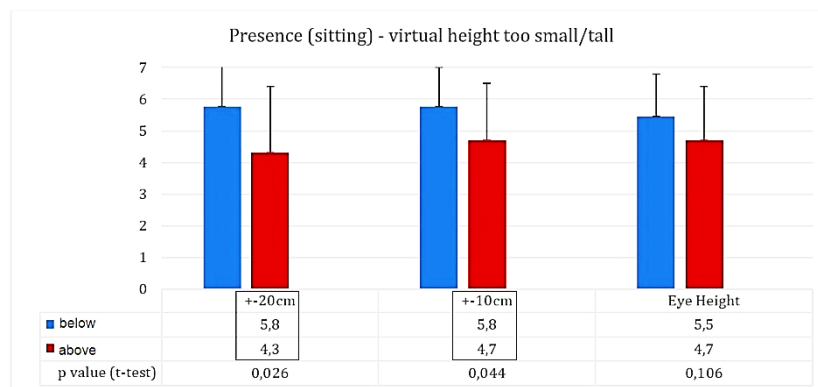


Abbildung 7-59: Für die allgemeine Presence sind die Mittelwerte für die *below*-Gruppe signifikant höher, wenn die Kamera 20 cm oder 10 cm von der Augenhöhe abweicht (Likert-Skala von 0 bis 7).

Experience

46.15 % der Teilnehmer beklagten sich über das erste Video (20 cm niedriger oder höher als die tatsächliche Höhe), da die Proportionen nicht stimmten und nur 20 % über das zweite Video (10 cm niedriger oder höher als die tatsächliche Höhe). Wenn man die beiden Gruppen vergleicht, gibt es einen Unterschied. 61.5 % der Zuschauer hatten ein seltsames Gefühl, das Video 20 cm zu hoch anzuschauen, während nur die Hälfte dieses Gefühl für das Video hatte, das 20 cm zu niedrig aufgenommen wurde. (Abbildung 7-60).

Die meisten Teilnehmer, die zuerst das Video über Augenhöhe sahen, bevorzugten das Video, das ihrer Größe entsprach (61.5 %). In der anderen Gruppe gaben jedoch nur 23.1 % an, dass ihnen das Video mit der tatsächlichen Größe am besten gefiel. Für sie war der Unterschied weniger wichtig.

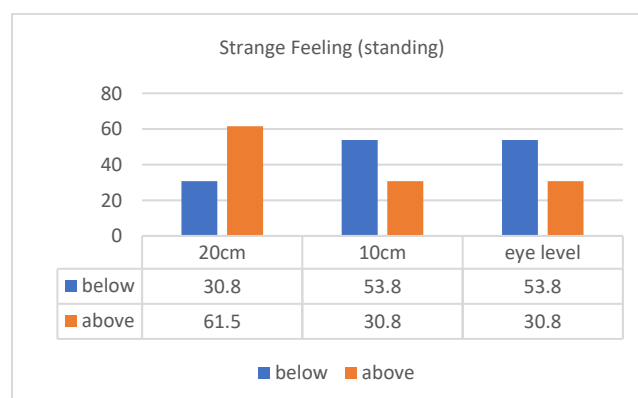


Abbildung 7-60: Das Diagramm zeigt Unterschiede im Zuschauererlebnis (in Prozent). Kamera und Augenhöhe unterscheiden sich um 20 cm bzw. 10 cm. In Situationen, in denen die Kameraposition höher als die tatsächliche Augenhöhe ist, geben mehr Betrachter seltsame Gefühle an als bei niedrigen Positionen. Es gibt auch einen Unterschied für die beiden Gruppen im Baseline-Fall.

Vergleich

84.6 % der Teilnehmer der *above*-Gruppe erkannten das in ihrer Größe aufgenommene Video, 30.8 % in der *below*-Gruppe (Tabelle 7-7).

		+20 cm	+10 cm	eye level
Welches Video würden Sie bevorzugen?	<i>below</i>	38.5 %	15.4 %	23.1 %
	<i>above</i>	15.4 %	23.1 %	61.5 %
Welches Video entspricht Ihrer eigenen Augenhöhe?	<i>below</i>	23.1 %	38.5 %	30.8 %
	<i>above</i>	0 %	15.4 %	84.6 %

Tabelle 7-7: Oben: Mehr Zuschauer bevorzugten das Video in ihrer eigenen Augenhöhe, wenn die Kameraposition zu hoch war. Unten: Mehr Zuschauer identifizierten das Video in der eigenen Augenhöhe, wenn das andere zu hoch war.

7.2.3.4 Diskussion und Einschränkungen

Die Ergebnisse dieser Studie werden ausführlich in Abschnitt 3.2 diskutiert und fließen in die Guidelines für Kameraführung in Abschnitt 3.5 ein.

In den Studien wurden keine stereoskopischen Videos verwendet, da derzeit die meisten CVR-Videos monoskopisch produziert werden. Es kommen jedoch immer mehr Kameras mit stereoskopischen Funktionen auf den Markt. Die Verwendung von stereoskopischen Bildern kann zu anderen Ergebnissen führen und benötigt weitere Untersuchungen.

Unsere Studien decken nur einige mögliche Videoinhalte ab. Es gab zwei verschiedene Szenarien und der Inhalt könnte die Ergebnisse beeinflusst haben. Für unsere Untersuchungen verwendeten wir unter anderem eine Szene an einem Tisch in einem Café, an dem die Protagonisten dem Betrachter sehr nahe sind. Wogegen in der stehenden Fußballszene die Hauptfigur weiter weg ist. Für andere Szenarien sind weitere Untersuchungen erforderlich, insbesondere ob Entfernungen zu Objekten und Personen die Akzeptanz unterschiedlicher Augenhöhen beeinflussen. Es ist möglich, dass nicht nur die Höhe der Kamera die Wahrnehmung der unterschiedlichen Höhen von Kamera und Augen beeinflusst. Auch Entfernungen, Bewegungen und Affordances sollten berücksichtigt werden. Da eine Sitzhaltung der Stehhaltung vorgezogen wurde und aus verschiedenen Gründen auch bequemer ist, sollten sich zukünftige Forschungen auf diesen Fall konzentrieren.

Einige der Teilnehmer hatten vorher noch nie einen Film über ein HMD gesehen, was die Ergebnisse möglicherweise beeinflusst hat. Aber auch für andere Nutzer kann sich das Betrachtungsverhalten im Laufe der Zeit ändern. Ungewöhnliche Ansichten können bei zunehmender Erfahrung akzeptabel werden.

7.2.4 Studie: Diskrepanzen in der Körperhaltung

Die Studie ist Teil der Bachelorarbeit von Boris Kegeles (2017): Influences of Camera Position to Attention and Presence in CVR.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Die Ergebnisse dieser Studie wurden vorab veröffentlicht in:

- Rothe, S., Kegeles, B., Hussmann, H., 2019. Camera Heights in Cinematic Virtual Reality, in: Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '19. ACM Press, New York, New York, USA
- Rothe, S., Kegeles, B., Hussmann, H., 2019. Camera Heights in Cinematic Virtual Reality, in: Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '19. ACM Press, New York, New York, USA

Ziel dieser Studie war es herauszufinden, wie unterschiedlich Körperhaltungen (sitzend/stehend) das Zuschauererlebnis beeinflussen.

Die Studie ist in zwei Teile gegliedert:

- Sitzend wurden Videos angeschaut, die in Steh-Höhe aufgenommen wurden (Abbildung 7-61, links).
- Stehend wurden Videos angeschaut, die in Sitz-Höhe aufgenommen wurden (Abbildung 7-61, rechts).

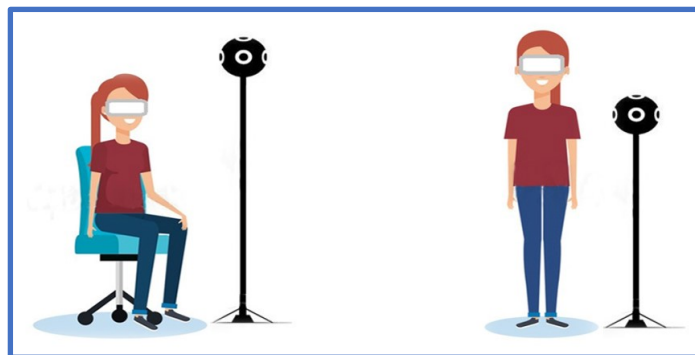


Abbildung 7-61: Sitzend schauten die Teilnehmer einen Film an, der in Stehhöhe aufgenommen wurde, und stehend einen Film in Sitzhöhe.

7.2.4.1 Material

Das Setup war das gleiche wie in **Studie 7.2.3**. Es wurde erneut ein Samsung Gear VR-Headset genutzt und das Videomaterial aus **Studie 7.2.3** wiederverwendet.

7.2.4.2 Teilnehmer und Prozedur

20 Teilnehmer (8 weiblich, 12 männlich, Alter: MW=32.15 SD = 11.3, Körpergröße: MW=179.4 cm SD = 9.75) sahen den Film über HMD. 65 % der Teilnehmer hatten VR-Erfahrungen. Personen, die an **Studie 7.2.3** teilgenommen hatten, waren für diese Studie nicht zugelassen.

Jeder Teilnehmer schaute sich zwei Videos an:

- Das in der Sitzhöhe aufgezeichnete Café-Video wurde stehend betrachtet. Die Kamerahöhe betrug 120 cm.
- Das in Stehhöhe aufgezeichnete Fußballvideo wurde in sitzender Haltung angesehen. Die Kamerahöhe betrug 170 cm.

Es wurde ein within-subject Testdesign ausgewählt. Die Reihenfolge der Videos wurde permutiert. Der Fragebogen war ähnlich dem der **Studie 7.2.3**: Die Fragen zur Person, der Presence und der Simulator-Sickness waren die gleichen. Zusätzlich wurden die Teilnehmer nach beiden Videos gefragt, ob sie ihre Pose beim Anschauen (vom Sitzen zum Stehen oder umgekehrt) ändern wollten.

7.2.4.3 Ergebnisse

Presence und Simulator-Sickness

Es gab keine signifikanten Unterschiede in Bezug auf Simulator-Sickness und Presence.

Experience

Mehr Teilnehmer gaben ein seltsames Gefühl an, im Stehen nicht die richtige Größe zu haben (Abbildung 7-62, links). Die Mehrheit hatte das Bedürfnis sich hinzusetzen, um das Video anzuschauen, das auf einer Sitzhöhe aufgenommen wurde (Abbildung 7-62, Mitte). Nur wenige Teilnehmer wollten aufstehen, um sich ein Video anzusehen, das für eine stehende Position aufgenommen wurde. Bei der Frage, welches der Videos schwieriger zu sehen ist, entschieden sich 55 % für das Sitzvideo und 30 % für das Stehvideo (Abbildung 7-62, rechts).

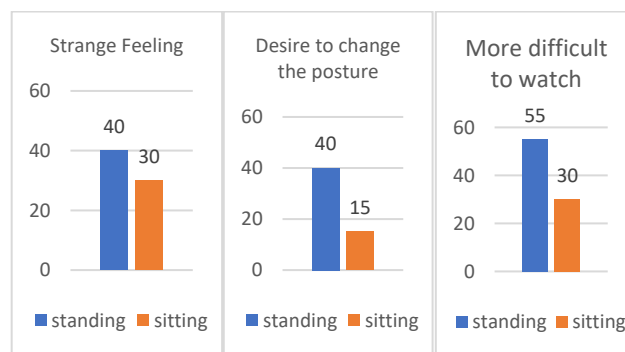


Abbildung 7-62: Stehende Personen sahen sich ein Video in Augenhöhe einer sitzenden Person an und umgekehrt. links: Anzahl der Teilnehmer (in Prozent) mit einem seltsamen Gefühl beim Ansehen des Videos. Mitte: Anzahl der Teilnehmer (in Prozent), die den Wunsch hatten, ihre Haltung zu ändern. rechts: Für mehr Teilnehmer war es schwierig, das Sitzvideo in stehender Haltung als das Stehvideo in sitzender Haltung (in Prozent) anzusehen.

Diesmal zeigte der exakte Fisher-Test keinen statistisch signifikanten Unterschied darin, „seltsame Gefühle“ ($p = 0.21$) zu haben, wenn der Film in der falschen Höhe angesehen wurde. Der Wunsch, die richtige Haltung einzunehmen, war jedoch für die Stehenden signifikant höher (exakter Fisher-Test, $p = 0.06$).

Alle Teilnehmer hatten die Möglichkeit, allgemeine Kommentare zur CVR-Experience abzugeben. P2 erwähnte nach dem Anschauen des Sitzvideos im Stehen: „Im Vergleich zum anderen Video fühlte ich mich kleiner, aber nicht unverhältnismäßig (im Gegensatz zum Riesengefühl).“ Für P13 und P11 war es seltsam, das Gefühl zu haben, mit Fremden auf einem Tisch zu sitzen, und P9 wollte den Tisch fühlen. Andere Teilnehmer erwähnten die ungewöhnliche Position (P5, P6, P8). In Bezug auf das Stehvideo in der Sitzhaltung betrafen alle Bemerkungen die ungeeignete Ansicht (P2, P4, P9, P10, P12, P14). P2 erwähnte ein „seltsames Gefühl im Magen“ und P14 eine verzerrte Darstellung.

7.2.4.4 Diskussion und Einschränkungen

Die Ergebnisse dieser Studie werden ausführlich in Abschnitt 3.2 diskutiert und fließen in die Guidelines für Kameraführung in Abschnitt 3.5 ein. In dieser Studie wurde das gleiche Material wie in **Studie 7.2.3** genutzt und die Einschränkungen stimmen mit denen in Abschnitt 7.2.3.4 überein.

7.2.5 Studie: Kamerabewegungen

Die Studie ist Teil der Bachelorarbeit von Nicole Gora (2019): Cuts and Camera Movements in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit der Studentin die Idee entwickelt, die sie implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir von der Studentin durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Ziel dieser Studie war es, herauszufinden, ob Kamerabewegungen vom Zuschauer akzeptiert werden, auch wenn er sich in der realen Welt nicht bewegt und damit die visuell wahrgenommene Bewegung nicht mit den Informationen der Sinnesorgane übereinstimmt. Dazu wurden drei Fälle miteinander verglichen:

cut

Die verschiedenen Positionen wurden allein durch Schnitte realisiert. Nach dem Schnitt befanden sich die Zuschauer an einer anderen Position. Die Blickrichtung wurde nicht geändert. Die Teilnehmer schauten in die gleiche Richtung wie vor dem Schnitt, konnten sich frei umsehen und dadurch die Umgebung entdecken. Es gab insgesamt 6 verschiedene Kamerapositionen (Abbildung 7-63, oben). An jeder der Positionen verweilte die Kamera ungefähr 28 s. Die Übergänge waren harte Schnitte (ohne Überblendung).

move (Kamerafahrt)

Eine gleichmäßige, geradlinige Kamerafahrt ermöglichte die Erkundung der Umgebung und das Beobachten der Charaktere (Abbildung 7-63, Mitte). In einer informellen Pilotstudie wurden verschiedene Geschwindigkeiten der Kamerabewegung getestet. Ursprünglich war geplant, dass alle Filme die gleiche Länge haben und die Zeit, die bei der Cutting-Methode an den einzelnen Positionen verbracht wurde, für die Bewegung der Kamerafahrt genutzt wird. Die daraus resultierende Geschwindigkeit wirkte auf dem Desktop angenehm, erwies sich im HMD aber als zu schnell und wurde verringert.

mix (cut und move)

Bei dieser Methode wurden die beiden ersten Methoden gemeinsam angewendet. Ein Teil des Films wurde über Schnitte realisiert. In der Mitte gab es eine Kamerafahrt (Abbildung 7-63, unten).

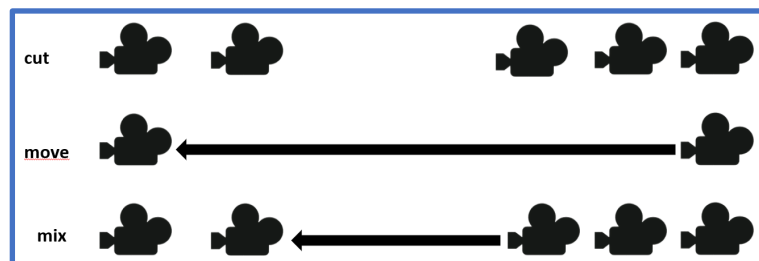


Abbildung 7-63: Darstellung der untersuchten Techniken: cut (oben), move (Mitte), mix (unten).

7.2.5.1 Material

Um möglichst sanfte Kamerabewegungen zu erhalten, entschieden wir uns in der Studie, eine Animation zu nutzen. Dafür wurde im Maya Autodesk eine 3D-Umgebung erzeugt und in Unity importiert. Die Umgebung war groß genug, um auch längere Kamerafahrten zu ermöglichen. Um die 3D-Umgebung mit sich bewegendenden Charakteren zu ergänzen, wurde Adobe Mixamo genutzt. Die Kamerabewegungen wurden in Unity 3D implementiert. Die Filme hatten eine Länge zwischen 2:48 min und 3:42 min und eine sehr einfache Story, damit der Inhalt nicht die Studienergebnisse beeinflusst: in der 3D-Umgebung tanzten fünf Charaktere zur gleichen Musik. Der Film ermöglichte eine Reise durch diese Landschaft, vorbei an all den tanzenden Personen und endete auf einer kleinen Insel mit einer Mühle. Alle drei Filme folgten der gleichen Storyline.

Während der Studie nutzten die Teilnehmer eine HTC Vive und Kopfhörer.

7.2.5.2 Teilnehmer und Prozedur

Es nahmen 24 Personen (11 weiblich, 13 männlich) im Alter von 18 bis 29 Jahren an der Studie teil. 22 von ihnen hatten bereits VR-Erfahrungen, aber keiner von ihnen nutzte regelmäßig ein HMD.

Es wurde ein within-subject Testdesign verwendet und jeder Teilnehmer sah den Film mit allen drei Methoden. Die Reihenfolge der Methoden wurde durch Permutationen ausbalanciert. Zu Beginn der Studie wurde jedem eine Testumgebung gezeigt, die die 3D-Umgebung der Studie zeigte, jedoch ohne die animierten Charaktere. Die Teilnehmer wurden aufgefordert sich umzuschauen, um den Novelty-Effekt und die intensivere Inspektion der Umgebung im ersten Teil der Studie zu reduzieren. Die folgenden abhängigen Variablen wurden per Fragebogen ermittelt: Presence, Simulator-Sickness, Usability und User Experience. Der Fragebogen endete mit einem Vergleich der Methoden.

7.2.5.3 Ergebnisse

Presence

Die Kamerafahrt (*moving*) erzielte die höchsten Scores für die Presence, gefolgt von der *cutting*-Methode. Der Mix aus beiden schnitt am schlechtesten ab. Allgemeine Presence: $MW_{cut}=5.2$ $SD_{cut}=1.38$ $MW_{move}=5.3$ $SD_{move}=1.16$ $MW_{mix}=4.67$ $SD_{mix}=1.2$ $p_{cut/mix}=0.039$ $p_{move/mix}=0.036$ (Abbildung 7-64).

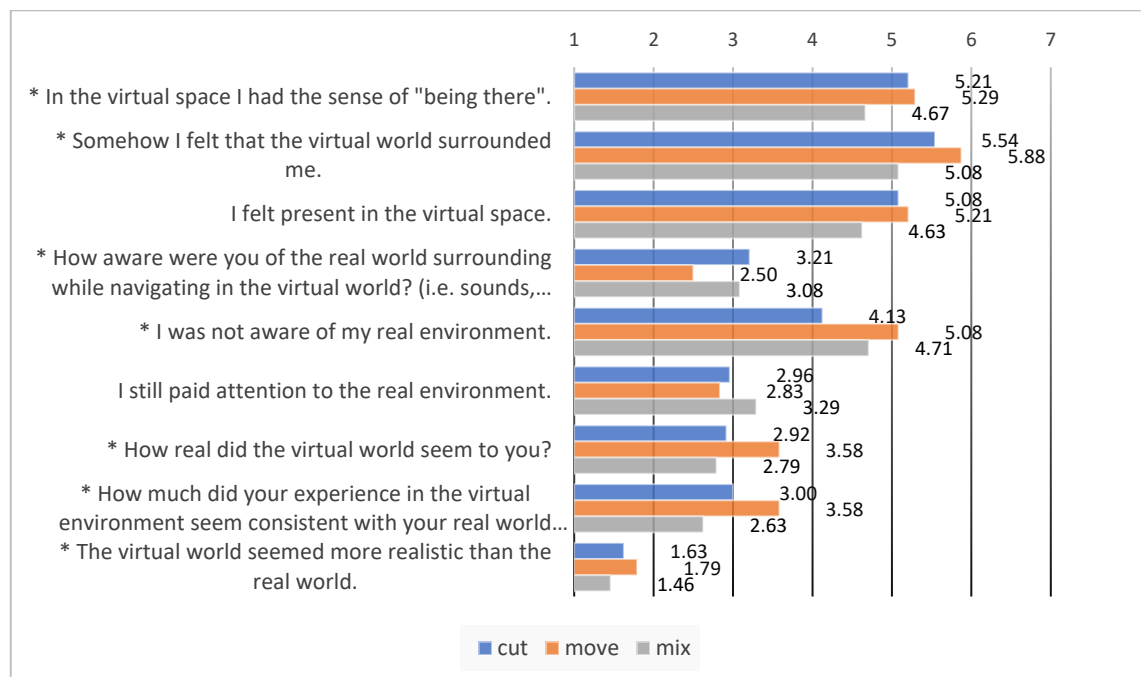


Abbildung 7-64: Mittelwerte der Presence-Scores (Likert-Skala von 1 bis 7). Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

Simulator-Sickness

Auch bezüglich der Simulator-Sickness schnitt die *mixed*-Methode am schlechtesten ab. Die anderen beiden Methoden unterschieden sich nicht signifikant (Abbildung 7-65).

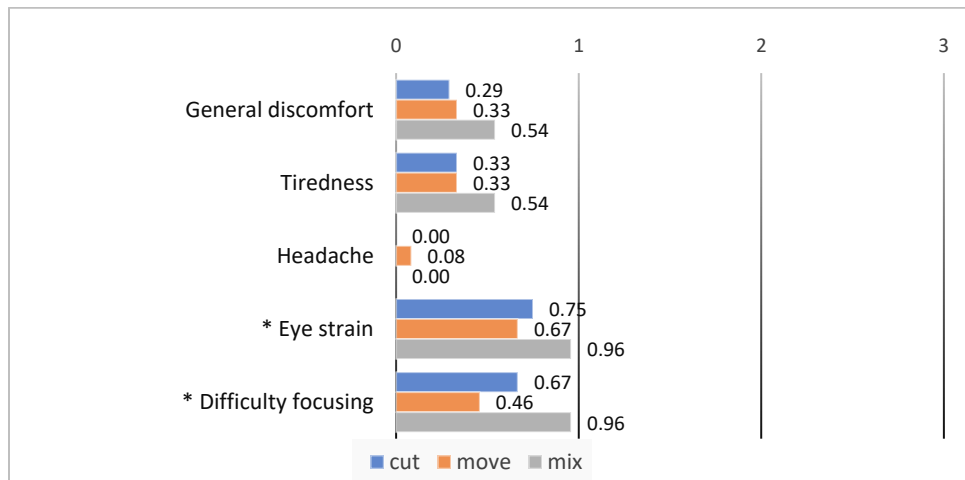


Abbildung 7-65: Mittelwerte für Scores des SSQ (Likert-Skala von 0 bis 3). Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

Usability

Die *moving*-Methode erzielte die besten Ergebnisse in den Usability-Scores. Die Unterschiede zwischen *moving*- und *mixed*-Methode sind für alle Items signifikant. Aber auch die *cutting*-Methode schnitt in mehreren Punkten schlechter als die *moving*-Methode ab (Abbildung 7-66).

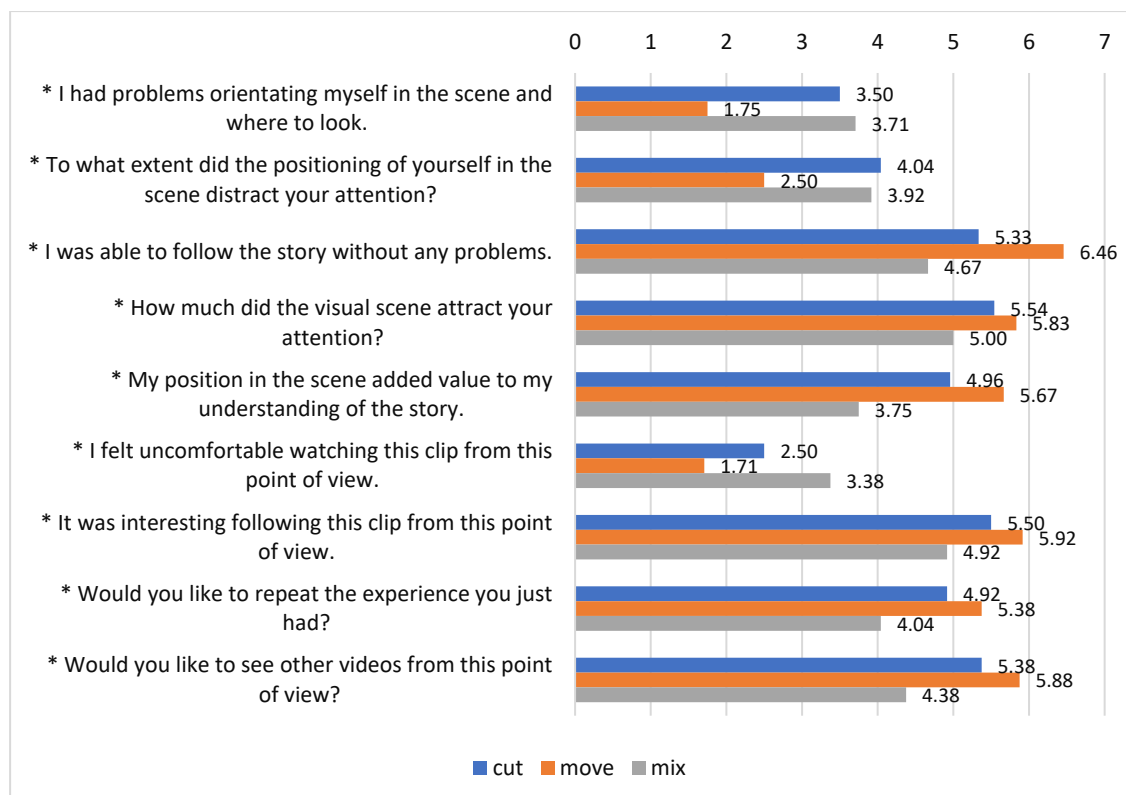


Abbildung 7-66: Mittelwerte für die Usability-Items (Likert-Skala von 1 bis 7). Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

User Experience

Auch bezüglich der User Experience erzielte die moving-Methode die besten Ergebnisse. Abbildung 7-67 zeigt die Werte für die Items des UX-Fragebogens, die signifikante Unterschiede aufwiesen.

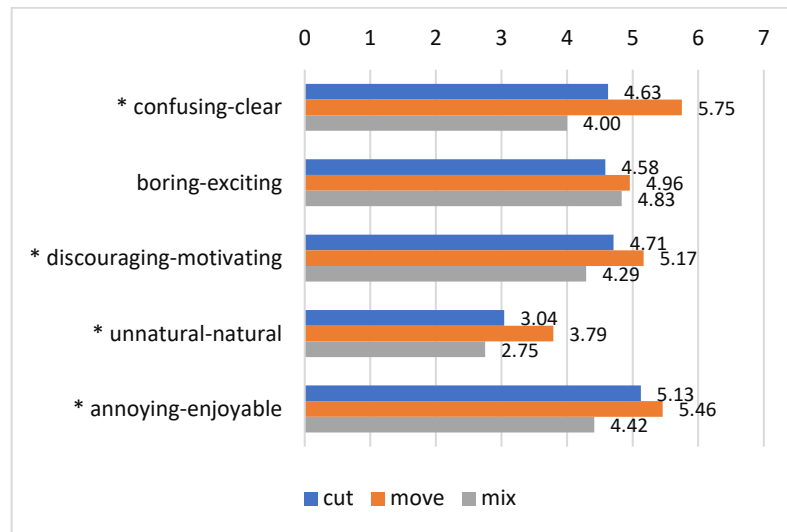


Abbildung 7-67: Mittelwerte für Items des UX-Fragebogens (Likert-Skala von 1 bis 7). Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

7.2.5.4 Diskussion

Die Ergebnisse der Nutzerstudie zeigen, dass flüssige Kamerabewegungen von den Zuschauern akzeptiert werden, auch wenn die Bewegung in der virtuellen Welt nicht mit der Bewegung in der realen Welt übereinstimmt. Ortswechsel durch Schnitte führten zu mehr Desorientierung als die Bewegung mit der Kamera. Am schlechtesten schnitt der Mix aus Kamerafahrt und Schnitten ab. Alle erzielten Ergebnisse hängen sowohl von den gewählten Parametern, wie Geschwindigkeit der Kamerabewegung, Ausrichtung der Schnitte, wie auch vom verwendeten Filmmaterial ab.

Das Sehverhalten der Zuschauer kann sich mit der Zeit ändern, so dass es möglich ist, dass Zuschauer sich mit der Zeit an Schnitte gewöhnen. Es bedarf allerdings weiterer Untersuchungen, wie diese Schnitte ausgeführt werden sollten, um dem Zuschauer eine angenehme Benutzererfahrung vermitteln zu können. Die wird in den Studien 7.3.1 und 7.3.2 analysiert.

7.3 Nutzerstudien: Montage

Um das Spaceline-Konzept aus Kapitel 4 umzusetzen, werden Erfahrungen benötigt, wie sich Richtungswechsel bei einer Transition zwischen Spaces auswirken (**Studie 7.3.1**). Um dem Zuschauer die benötigte Zeit in einem Space zur Verfügung zu stellen wird eine Wait-Loop untersucht (**Studie 7.3.2**). Einige on- und off-screen Indikatoren zum Anzeigen von Rols werden vorgestellt und miteinander verglichen (**Studien 7.3.3 und 7.3.4**). Da die Aktivierung der Regionen für den Zuschauer problemlos möglich sein soll, ohne das Zuschauererlebnis zu beeinträchtigen, wurden verschiedene Interaktionsmethoden miteinander verglichen (Studie 7.3.5). Für die Auseinandersetzung mit der Achsensprungproblematik, wurden verschiedene Positionen und Richtungen der Kamera in aufeinanderfolgenden Filmsequenzen untersucht (Studie 7.3.6).

In unserem Spaceline-Konzept wurde der Begriff Transition für den Space-Wechsel eingeführt. Da wir während der Studien noch den Begriff Schnitt (engl.: cut) verwendeten, werden in diesem Abschnitt beide Begriffe synonym verwendet.

7.3.1 Studie: Richtungswechsel bei Transitions

Die Studie ist Teil der Bachelorarbeit von Dominik Beyer (2019): *Camera Positions in Cinematic Virtual Reality*.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

In dieser Studie werden *co-located* Schnitte untersucht, d. h. die Kamerapositionen befinden sich vor und nach den Schnitten im gleichen Raum. Die drei implementierten Methoden unterscheiden sich in der Ausrichtung des Blickes nach dem Schnitt.

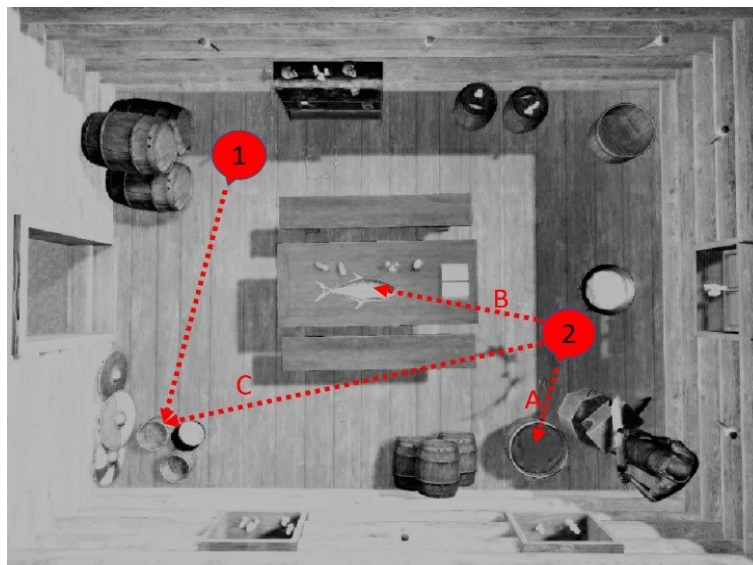


Abbildung 7-68: Szene der Studie mit einer Kameraposition vor (1) und nach (2) dem Schnitt. Die Pfeile zeigen die Blickrichtungen. Die Richtung nach dem Schnitt hängt von der verwendeten Methode ab: *keepDir* (A), *toRoI* (B), *keepFocus* (C).

keepDir

Die Blickrichtung zwischen den Schnitten wird beibehalten - lediglich die Position ändert sich. Abbildung 7-68 zeigt ein Beispiel für Positionen vor (1) und nach (2) dem Schnitt. Die Richtungen werden durch Pfeile angezeigt. Nach dem Schnitt hat sich die Blickrichtung nicht geändert (Abbildung 7-68, Pfeil A).

toRoI

Die Blickrichtung zeigt nach dem Schnitt auf eine neue RoI (Abbildung 7-68, Pfeil B).

keepFocus

Die Kamera ändert die Position, der Fokus des Betrachters bleibt jedoch unberührt. Das Objekt, welches sich vor dem Schnitt im Sichtfeld befand, ist auch danach im Focus (Abbildung 7-68, Pfeil C).

7.3.1.1 Material

Da wir hinsichtlich der Kamerapositionen flexibel sein wollten, wurde die Videoszene nicht in der realen Welt aufgenommen. Es wurde ein 3D-Modell einer historischen Hütte erstellt, das mehrere RoIs enthält, die durch Lichter oder Bewegungen (z. B. einen arbeitenden Schmied und brennende Fackeln) auffallen. Im Modell können die Kamerapositionen leicht geändert werden. Die Methoden wurden in Unity 2018.3

implementiert und eine HTC Vive zum Betrachten der Anwendung verwendet. Die Zuschauer konnten sich nicht im Raum bewegen, lediglich den Kopf drehen.

Der Film beinhaltete keine Story, da wir uns für die Wirkungen der verschiedenen Schnitte interessierten, die von der Geschichte nicht beeinflusst werden sollten. Nach 20 s änderte die Kamera die Position und je nach Methode auch die Richtung. Dabei wurden fünf verschiedene Kamerapositionen eingenommen.

7.3.1.2 Teilnehmer und Prozedur

26 Personen (11 weiblich, 15 männlich, Alter: MW = 27.7, SD = 12.8) nahmen an der Studie teil. Für 10 Teilnehmer war es das erste VR-Erlebnis. Es wurde ein within-subject Testdesign verwendet, und jeder Teilnehmer sah die Szene mit allen drei Methoden. Die Reihenfolge der Methoden wurde durch ein lateinisches Quadrat ausgeglichen.

Unter Verwendung etablierter Fragebögen wurden die Presence (Schubert u. a., 2002) und die Simulator-Sickness (SSQ) (Kennedy u. a., 1993) gemessen. Zusätzlich wurden spezielle Fragen zur Benutzererfahrung gestellt und nach jeder Methode ein kurzes informelles Interview geführt. Nachdem die Teilnehmer alle Methoden gesehen hatten, wurden einige abschließende Fragen zum Vergleich beantwortet.

7.3.1.3 Ergebnisse

Der Shapiro-Wilk-Test zeigte, dass die Normalverteilung der Daten nicht gegeben ist. Daher wurde der Friedman-Test angewendet, um signifikante Unterschiede zwischen den Bedingungen zu finden (α -Level von 0,05). Falls dieser Test Signifikanz zeigte, wurden zusätzlich post-hoc paarweise Wilcoxon-Signed-Ranks-Tests ausgeführt.

Presence

Der Friedman-Test zeigte keine signifikanten Unterschiede zwischen den drei Methoden ($p=0.9$).

Simulator-Sickness

Die Werte für die Simulator-Sickness waren für alle Items sehr niedrig (zwischen 0.07 und 0.46), und die Unterschiede zwischen den Methoden ($p=0.16$) nicht signifikant.

Experience

Es gab auch keine signifikanten Unterschiede in den Antworten auf die Experience-Fragen. Um einen allgemeinen Überblick über die Ergebnisse zu geben, sind die Mittelwerte in Abbildung 7-69 dargestellt. Sie zeigen, dass die Teilnehmer nach einem Schnitt kaum Probleme hatten, sich zu orientieren (E1, E2, E3). Auch wenn die Schnitte zu einer besseren Übersicht beitragen (E5) und die Konzentration auf Details erleichtern (E7), lenken sie ein wenig vom Inhalt ab (E6).

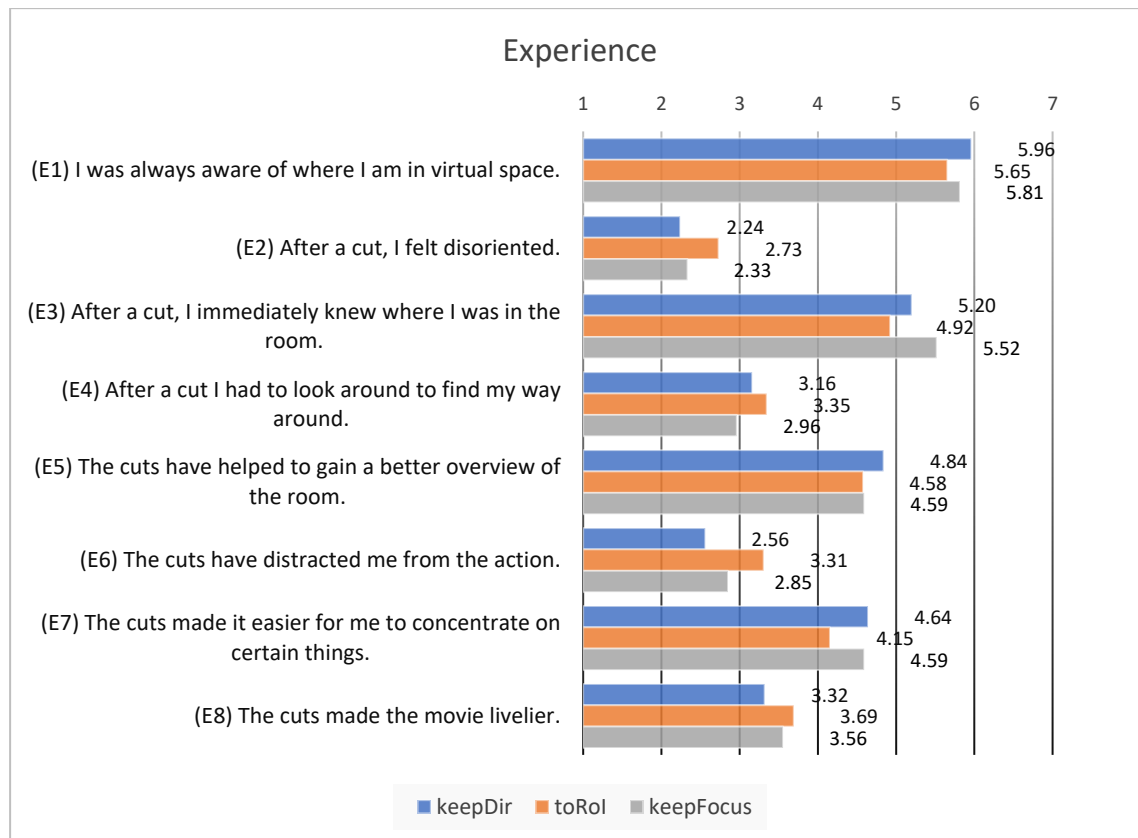


Abbildung 7-69: Experience-Fragen und Mittelwerte für jede Methode (Likert-Skala von 1 bis 7).

Vergleich

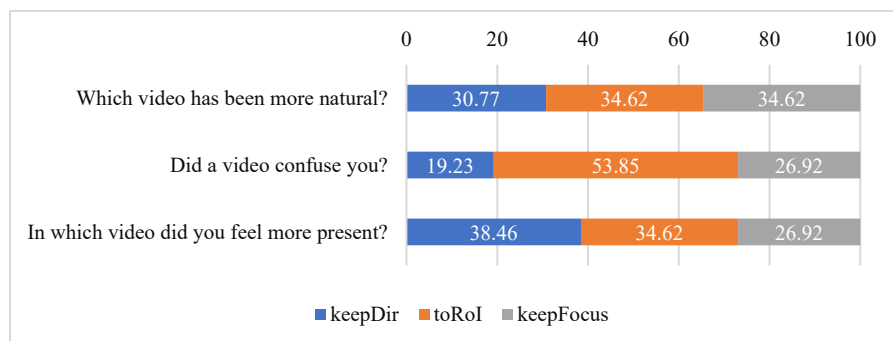


Abbildung 7-70: Ergebnisse der Vergleichsfragen (in Prozent).

Auch beim direkten Vergleich der Methoden wurden keine signifikanten Unterschiede festgestellt. Im kurzen informellen Interview gaben die meisten Teilnehmer an, dass sie keine Unterschiede zwischen den Testfällen erkannten. Allerdings waren die meisten Teilnehmer (53.85 %) von der *toRoI*-Methode verwirrt (Abbildung 7-70).

7.3.1.4 Diskussion

Bei den meisten Fragen konnten wir keine signifikanten Unterschiede zwischen den Methoden feststellen. Dies stimmt mit dem Bericht der Teilnehmer überein, Unterschiede zwischen den Testfällen nicht wahrgenommen zu haben.

Der einzige Unterschied war das Gefühl der Desorientierung. Mehr als die Hälfte der Teilnehmer war von der *toRoI*-Methode verwirrt. Im Gegensatz zur **Studie 7.3.6** gab es in dieser Studie keine Story. Darüber hinaus befanden sich einige der Rols weiter von der 0°-Horizontlinie entfernt. Objekte lagen auf dem

Tisch, der Betrachter musste darauf herabsehen. Mit gerichteten Schnitten kann man nur die horizontale Richtung ändern, Änderungen in der vertikalen Richtung würden die Welt „kippen“ lassen. Unser Ziel war es zu verstehen, ob sich Zuschauer nach unterschiedlichen Schnitten im Raum noch orientieren können. Es scheint, dass in Fällen, in denen die Rols nicht offensichtlich, für die Story nicht relevant oder weiter von der horizontalen 0°-Linie entfernt sind, der *toRol*-Schnitt den Betrachter verwirrt. In **Studie 7.3.6** gab es diese Desorientierung nicht. Dort waren die Zuschauer nicht darauf fokussiert, den Raum zu inspizieren. Stattdessen lag die Konzentration auf dem Gespräch, welches durch den *toRol*-Schnitt besser verfolgt werden konnte.

Mit diesen Methoden können die Teilnehmer nicht direkt zu Rols geführt werden, die sich nicht auf ihrer aktuellen vertikalen Augenhöhe befinden. Daher sind in solchen Bereichen weitere Forschungsarbeiten erforderlich. Eine mögliche Methode wurde in der nächsten Studie untersucht.

7.3.2 Studie: Loops vor Transitions

Die Studie ist Teil der Bachelorarbeit von Evelyn Müller, (2018): Differences Between Cinematic VR and Traditional Movies.

Ich habe diese Arbeit betreut und gemeinsam mit der Studentin die Idee entwickelt, die sie implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir von der Studentin durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Um den Betrachter die Möglichkeit zu geben, sich in Ruhe in einer Szene umzuschauen und auf Rols aufmerksam zu werden, wurde die *wait*-Methode implementiert, die in dieser Studie mit der *toRol*-Methode verglichen wird.

toRol

Der Blick wird so ausgerichtet, dass die Rol nach dem Schnitt im Focus des Betrachters liegt (wie in **Studie 7.3.6**)

wait

Es werden vorab Rols in einer Szene definiert. Der Film wird fortgesetzt, wenn der Betrachter alle Rols gesehen hat. Ein Objekt wird als „gesehen“ markiert, wenn der Betrachter mindestens 2 s lang in dessen Richtung blickt. Nach 20 s wird das Video auch dann fortgesetzt, wenn nicht alle Rols gesehen wurden. Die Größe der Bereiche um die Objekte betrug ca. 35°.

Als Baseline wurden die Originalvideos ohne zusätzliche Lenkungsmethoden verwendet. In den verwendeten Videos wurden die Aufnahmen von den Filmemachern in der Regel so verknüpft, dass einem Bereich, der die Aufmerksamkeit der Betrachter auf sich zog, nach dem Schnitt eine Rol im neuen Space folgte.

7.3.2.1 Material

Es wurden drei professionelle Videos über Museen ausgewählt, die von der New York Times produziert wurden (New York Times, 2017a, 2017b, 2017c). Alle Videos waren von gleichem Typ (kurze Dokumentationen) und gleicher Länge (ca. 1:30 min). Die Videos enthalten bereits Schnitte und zeigen Bildinformationen im gesamten 360°-Bereich. Die Einstellungen zwischen den Schnitten haben unterschiedliche Längen zwischen 7 s und 20 s (MW = 11.6 s, SD = 4.5). Die Schnittfrequenz war bei einem der Videos zum Teil relativ hoch. In einer VR-Umgebung benötigt der Betrachter allerdings mehr Zeit für die Erkundung aller Richtungen, und der Betrachter hat möglicherweise Befürchtungen, etwas zu verpassen. Das kann das VR-Erlebnis weniger angenehm machen.

Da wir die typischen häuslichen Bedingungen untersuchen wollten, verwendeten wir ein Google-Cardboard mit einem Motorola Moto G5-Smartphone. Die Methoden wurden in Unity 2017 implementiert.

7.3.2.2 Teilnehmer und Prozedur

33 Personen (15 weiblich, 18 männlich, Alter: MW = 27, SD = 11.85) nahmen an der Studie teil. Es wurde ein within-subject Testdesign verwendet und zu Testbeginn ein kurzes Eingewöhnungsvideo gezeigt. Danach schaute sich jeder Teilnehmer drei Filme an, zwei davon mit einer der beiden Methoden und eines als Baseline ohne eine der Methoden. Die Reihenfolge der Methoden und Filme wurden unter Verwendung eines Graeco-Latin-Quadrats ausbalanciert (Abbildung 7-71).

A1	B2	C3
B3	C1	A2
C2	A3	B1

Abbildung 7-71: Lateinisch-griechisches Quadrat für drei Methoden (A, B, C) und drei Filme (1,2,3). Die ersten Teilnehmer sahen sich Film 1 mit Methode A, Film 2 mit Methode B usw. an.

Zu Beginn der Studie beantwortete jeder Teilnehmer Fragen zu Alter, Geschlecht und VR-Erfahrung. Unter Verwendung etablierter Fragebögen wurden Presence (IPQ) (Schubert u. a., 2002) und Simulator-Sickness (SSQ) (Kennedy u. a., 1993) gemessen. Darüber hinaus bewerteten die Zuschauer einige Items zum Filmerlebnis. Nachdem die Teilnehmer alle Videos gesehen hatten, wurden abschließende Fragen zum Vergleich der Methoden beantwortet.

Die Headtracking-Daten der Teilnehmer wurden während der Studie aufgezeichnet und zusammen mit den Ergebnissen der Fragebögen analysiert.

7.3.2.3 Ergebnisse

Presence

Der Friedman-Test zeigte keine signifikanten Unterschiede zwischen den drei Methoden. Die Werte für die allgemeine Presence lagen bei: $p=0.9$, $MW_{base}=4.54$, $SD_{base}=1.82$, $MW_{roi}=4.57$, $SD_{roi}=1.92$, $MW_{wait}=4.93$, $SD_{wait}=1.68$.

Simulator-Sickness

Die Scores für Simulator-Sickness waren generell niedrig (0.09-0.88) und wir konnten keine signifikanten Unterschiede feststellen ($p=0.89$).

Experience

Die Experience-Fragen wurden auf einer 7-Punkte-Likert-Skala beantwortet. Bei vier der sieben Items stellten wir signifikante Unterschiede fest: Abbildung 7-72 und Tabelle 7-8 zeigen die Mittelwerte und Standardabweichungen der Scores, zusätzlich die zugehörigen p-Werte der Friedman-Tests und der paarweisen post-hoc-Tests.

	base	wait	toRol	pFriedman	pbase/toRol	pbase/wait	pwait/toRol
E1	4.21(1.78)	3.24(1.92)	3.72(1.86)	0.1	0.3	0.03	0.2
E4	3.70(1.69)	2.90(1.33)	3.27(1.28)	0.06	0.25	0.05	0.26
E6	3.70(1.91)	5.42(1.68)	3.89(1.93)	0.02	0.72	0.002	0.001
E7	2.94(1.95)	4.18(2.10)	3.03(1.59)	0.06	0.8	0.02	0.01

Tabelle 7-8: Mittelwerte, Standardabweichungen (in Klammern) und p-Werte der Experience-Items.

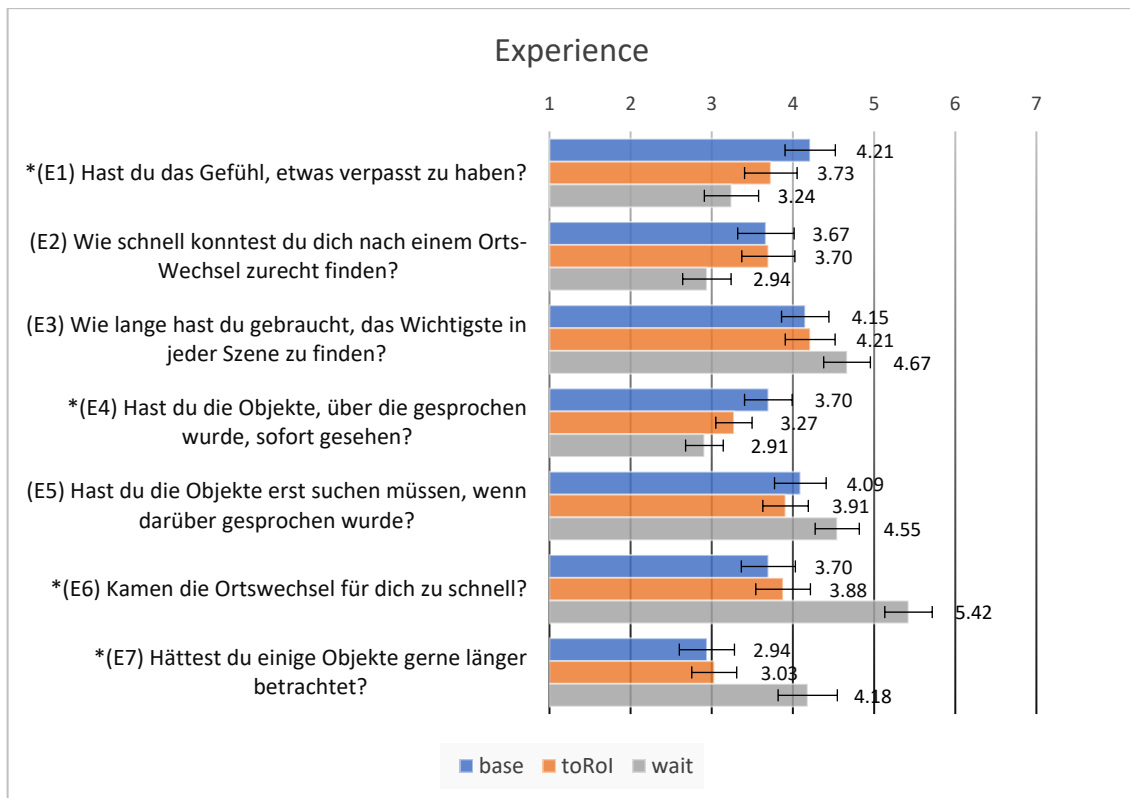


Abbildung 7-72: Mittelwerte und Standardfehler für die Experience-Fragen (Likert-Skala von 1 bis 7).

Vergleich

Mehr als die Hälfte der Teilnehmer (57.6 %) gab an, dass sie durch die *wait*-Methode das Gefühl hatten, weniger zu verpassen, bei der *toRol*-Methode waren dies 18.18 % und bei der Baseline 24.24 %. Bezüglich Frage (C2) waren die Unterschiede zwischen den Methoden geringer, was darauf zurückzuführen ist, dass der Film bei der *wait*-Methode nur fortgesetzt wurde, wenn alle Rols gesehen worden waren und die Zuschauer bei der *wait*-Methode die Rols selbst finden mussten. Den meisten Teilnehmer (57.6 %) gefiel die *wait*-Methode besser. (Abbildung 7-73).

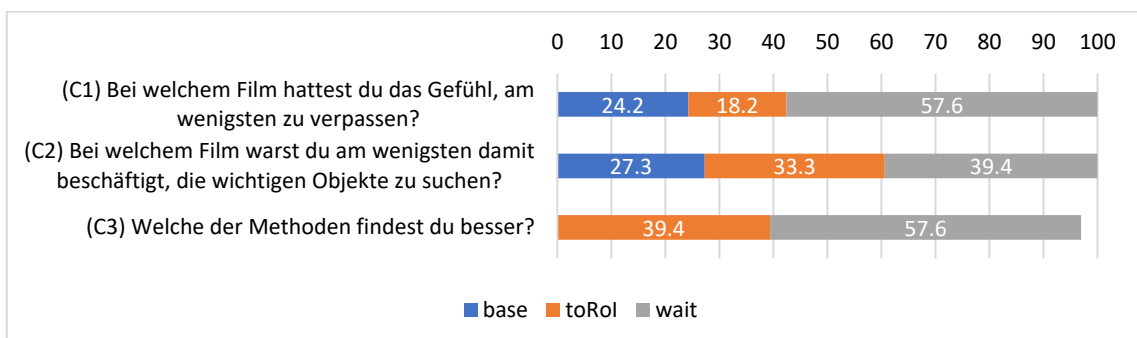


Abbildung 7-73: Ergebnisse des Vergleichs (in Prozent).

Analyse der Tracking-Daten

Die Tracking-Daten wurden auf zwei Arten mit dem CVR-Analyzer untersucht. Einerseits wurde das Hotspots-Diagramm (Abbildung 7-74) des CVR-Analyzers verwendet, um die Timecodes zu erkennen, auf die sich die Blicke der Teilnehmer konzentrierten. Andererseits wurden Heatmaps in der Nähe der Schnitte und für jeden relevanten Timecode untersucht (Rothe et al., 2018a).

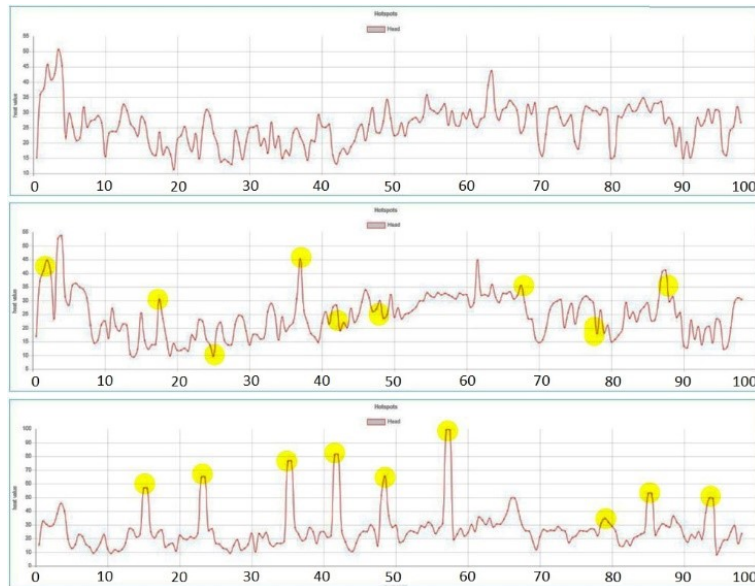


Abbildung 7-74: Hotspotdiagramm für die *Baseline* (oben), *toRoI* (Mitte), *wait* (unten). Die gelben Punkte zeigen die Positionen, an denen die Schnittmethoden angewendet wurden. Die x-Achse zeigt den Timecode im Film, die y-Achse die Heatvalues. Ein hoher y-Wert zeigt an, dass sich die Blickrichtungsdaten für diesen Timecode auf diesen Filmbereich konzentriert waren.

Beim Vergleich der Diagramme (Abbildung 7-74) fällt auf, dass die Graphen der *Baseline* und der *toRoI*-Methode sehr ähnlich sind. Häufig befinden sich die Hotspots am gleichen Timecode. Ein Grund dafür ist, dass wir kommerzielle Videos ausgewählt haben, bei denen die RoI mit der erwarteten Blickrichtung übereinstimmt. Auch in der *Baseline* erblickten die Zuschauer die meisten Rols. Wir erwarteten jedoch nach jedem *toRoI*-Schnitt, weitere herausragende Peaks, was nur bei wenigen Schnitten der Fall war. Bei der Untersuchung solcher Intervalle haben wir mehrere Gründe dafür gefunden: (1) Die Blicke der Teilnehmer waren nicht auf der vertikalen Höhe der Rols, z. B. schauten die Teilnehmer geradeaus, aber die Rol lag tiefer. (2) Einige RoI-Bereiche, die für die *toRoI*-Methode ausgewählt wurden, waren in vertikaler Richtung zu weit verteilt, um Heatmap Hotspots zu erzeugen.

7.3.2.4 Diskussion

Die Ergebnisse dieser Studie werden in Abschnitt 4.2.3.3. diskutiert und dort für das Spaceline-Konzept und die dazugehörigen Guidelines genutzt.

7.3.3 Studie: On-Screen Indikatoren: Vergleich von Cursor und Shape

Die Studie ist Teil der Masterarbeit von Harald Brunner (2018): Affordances for Interaction in Cinematic Virtual Reality via WebVR.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Um die Visualisierungen von on-screen Indikatoren auf ihre Wirksamkeit und ihren Einfluss auf die CVR-Experience zu testen wurden zwei Indikatoren implementiert:

Cursor

Der Cursor zeigt die Mitte des Displays an und ist außerhalb einer RoI sehr subtil. In einer RoI wird die Farbe des Cursors auffälliger und der Cursor sichtbarer (Abbildung 7-75, links).

Shape

Ein Shape hat die Form der RoI und ist sichtbar, sobald die RoI im Sichtfeld liegt. Die Fläche ist teiltransparent, so dass der Filminhalt gesehen werden kann (Abbildung 7-75, rechts).



Abbildung 7-75: Trifft der Cursor auf eine RoI wird sein Aussehen auffällig (Cursor, links). Ein Shape als Overlay zeigt die RoI (Shape, rechts).

7.3.3.1 Material

Das für die Studie verwendete Projekt besteht aus verschiedenen Szenen in einem keltischen Museumsdorf und entspricht der Story Struktur in Abbildung 7-76. Eine Szene zeigt das Dorf, die anderen das Innere verschiedener Hütten. In diesen Szenen erzählt ein Druide über das Leben der Kelten. Die RoIs der Dorfszene befinden sich auf den Hütten, die RoI der Hüttenszenen auf den jeweiligen Türen. Darüber hinaus enthalten die Szenen insgesamt sieben historische Textinformationen, die durch Aktivierung der zugehörigen Region sichtbar werden.

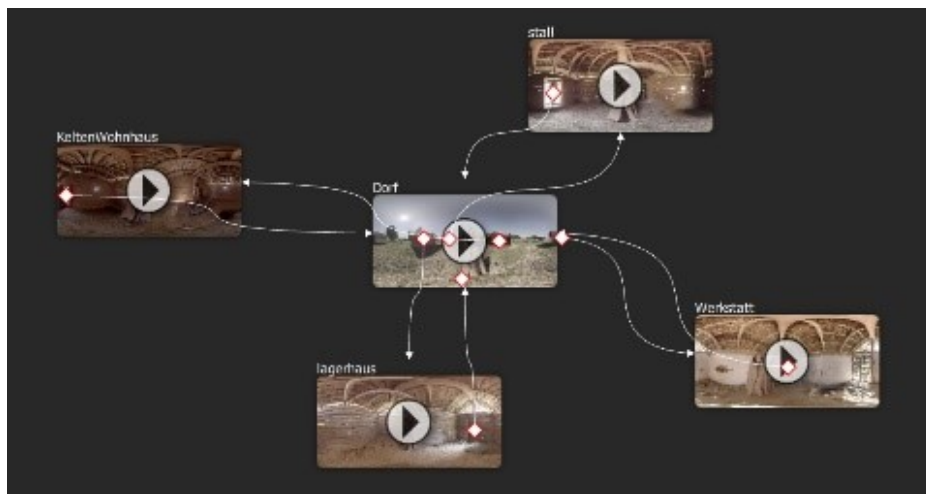


Abbildung 7-76: Testszene für die Visualisierung von on- und off-screen Indikatoren.

7.3.3.2 CVR-Editor

Für die beschriebene Untersuchung wurde ein Prototyp eines CVR-Editors entwickelt (Abschnitt 4.3.3.1). Mit dem Editor können verschiedene Spaces zu einem Spaceline-Konstrukt kombiniert werden. Dies kann durch die Definition von in- und out-Regionen sowie durch Definieren von on- und off-screen Indikatoren geschehen. Darüber hinaus können mit dem Tool weitere Interaktionen für *add*-Regionen (z. B. Text-Popups) definiert werden.

7.3.3.3 Teilnehmer und Prozedur

Von den 16 Teilnehmern (7 weiblich, 9 männlich, Alter: MW=22.2) hatten 87.5 % VR-Erfahrung. Zum Aktivieren von Regionen wurde der Oculus Controller verwendet. Es wurde eine Studie auf der Grundlage eines between-subject Testdesigns durchgeführt. Dazu wurden die Teilnehmer in zwei Gruppen eingeteilt und für jede Gruppe andere Indikatoren zum Anzeigen der Interaktionsmöglichkeiten genutzt. Die Teilnehmer saßen während der Studie auf einem Drehstuhl und nutzten ein Oculus Rift HMD.

In beiden Gruppen diente der *Cursor* als on-screen Indikator. Durch Drücken der *Controller*-Taste gelangte der Betrachter zur nächsten Einstellung, wenn die Rol anvisiert wurde. Für eine Gruppe war der Cursor der einzige verfügbare Indikator. Für die andere Testgruppe erhielten die Rols ein halbtransparentes Overlay (*Shape*) als zusätzlichen on-screen Indikator (Abbildung 7-75).

Die Indikatoren wurden den Teilnehmern zuvor nicht erklärt, sie sollten intuitiv verstanden werden.

Für die Studie wurden etablierte Fragebögen verwendet: Igroup Presence Questionnaire (IPQ) für die Presence (Schubert u. a., 2002) und Simulator-Sickness Questionnaire (SSQ) für Simulator-Sickness (Kennedy u. a., 1993). Außerdem wurden folgende User-Experience-Items abgefragt: störend, hilfreich, verständlich.

7.3.3.4 Ergebnisse

Presence und Simulator-Sickness

Wie aus Tabelle 7-9 ersichtlich, war die Presence für die Cursor-Bedingung signifikant höher.

	Realismus	Involvierung	Spatial	Gesamt
Cursor	4.88 (0.91)	5.56 (1.19)	5.22 (0.45)	5.26 (0.66)
Shape	3.92 (0.99)	4.53 (1.26)	4.59 (0.91)	4.41 (0.92)
p-Wert	0.03	0.06	0.049	0.03

Tabelle 7-9: Ergebnisse (Mittelwerte und Standardabweichungen in Klammern) für die IPQ-Skala (7-Punkte-Likert-Skala) und das Gesamtergebnis pro Gruppe. Die signifikant unterschiedlichen Werte (paarweiser t-Test, $p < 0.05$) sind fett markiert.

Es gab kaum Simulator-Sickness. Die Unterschiede zwischen den Gruppen waren diesbezüglich nicht signifikant. Auch die Bewertung der Belastung zeigte keine signifikanten Unterschiede zwischen den Gruppen.

User Experience

Der Cursor wurde kaum als störend empfunden, die Unterschiede zwischen Cursor und Shape waren bezüglich dieses Items nicht signifikant (Tabelle 7-10). Beide Indikatoren waren hilfreich und verständlich, wobei der Cursor als hilfreicher und die Shape-Bedingung als verständlicher eingestuft wurden. Die Teilnehmer der Shape-Gruppe gaben dem Shape-Indikator die volle Punktzahl für die Verständlichkeit.

	störend	hilfreich	verständlich
Cursor	1.25 (0.46)	6.50 (0.76)	6.13 (0.83)
Shape	2.25 (2.05)	5.38 (1.69)	7.00 (0)
p-Wert	0.1	0.049	0.01

Tabelle 7-10: Ergebnisse (Mittelwerte und Standardabweichungen in Klammern) für die User-Experience-Items (7-Punkte-Likert-Skala). Die signifikant unterschiedlichen Werte (paarweiser t-Test, $p < 0.05$) sind fett markiert.

Effektivität

Zusätzlich zum Fragebogen wurden die Aktionen der Zuschauer aufgezeichnet. Alle Teilnehmer verwendeten jede Region, um zu den anderen Szenen zu wechseln.

Es gab jedoch erhebliche Unterschiede bei der Nutzung der sieben Popup-Fenster. Im Durchschnitt wurden in der Cursor-Gruppe nur zwei der *add*-Regionen aktiviert. In der anderen Gruppe waren es vier.

7.3.3.5 Diskussion

Der in beiden Gruppen verwendete Cursor erwies sich als hilfreich und kaum störend. Es ist interessant, dass der Cursor im Fragebogen in der Kategorie „hilfreich“ die höchste Bewertung erhielt, aber dennoch zu den wenigsten Interaktionen führte. Dies liegt daran, dass die Betrachter aufgrund des Fehlens anderer Indikatoren nicht bemerkt haben, dass weitere Interaktionen möglich waren. Die zusätzliche Verwendung des Shapes führte zu einer Verdoppelung der Interaktionen. Das Ergebnis zeigt, dass in CVR neben dem Cursor weitere Indikatoren benötigt werden, wenn die RoI für die Story relevant ist.

Die vorgestellten Indikatoren sind Beispiele für die Anzeige von RoIs. Es gibt jedoch auch andere Möglichkeiten für on-screen Indikatoren, die es wert sind evaluiert zu werden. Die Indikatoren unterstützen die Interaktion des Betrachters.

Die Aktivierung der Regionen erfolgte in unserer Studie durch Drücken der Controller-Taste. In **Studie 7.3.5** wurden weitere Aktivierungstechniken getestet.

7.3.4 Studie: Off-Screen Indikatoren: Vergleich von Linien und Kantenspots

Die Studie ist Teil der Masterarbeit von Harald Brunner (2018): Affordances for Interaction in Cinematic Virtual Reality via WebVR.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Um die Visualisierungen von off-screen Indikatoren auf ihre Wirksamkeit und ihren Einfluss auf die CVR-Experience zu testen, wurden zwei Indikatoren implementiert:

Strahl

Die *Strahlen* haben ihren Ursprung in der RoI und führen zum Displaymittelpunkt. Der Gradient dieses Strahls ändert sich von blau-grün (im Ursprung) zu vollständig transparent (am Displaymittelpunkt). Richtung und Gradient ermöglichen es dem Betrachter, die Position der off-screen Region abzuschätzen (Abbildung 7-77, links). Die Anzeige verschwindet, wenn sich die RoI im FoV des Benutzers befindet.

Kantenspot

Kantenspots sind halbkreisförmige Objekte am Rand des Displays, die in Richtung der off-screen RoI weisen (Abbildung 7-77). Die Anzeige verschwindet, wenn die RoI im FoV des Benutzers liegt.

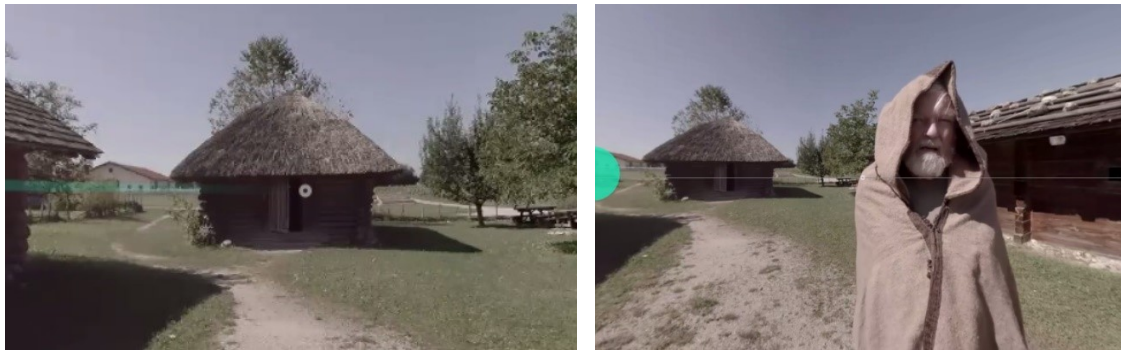


Abbildung 7-77: Zwei off-screen Indikatoren, links: Strahl, rechts: Kantenspot.

7.3.4.1 Teilnehmer und Prozedur

Von den 17 Teilnehmern (8 weiblich, 9 männlich, Alter MW=28.2) hatten 76.5 % VR-Erfahrung. Es wurde eine Studie auf der Grundlage eines between-subject Testdesigns durchgeführt, bei dem jede Gruppe das interaktive Filmprojekt mit verschiedenen Arten von Indikatoren testete. Die Teilnehmer wurden in 2 Gruppen eingeteilt, für jede Gruppe wurden andere off-screen Indikatoren zum Anzeigen der Interaktionsmöglichkeiten verwendet. Die Teilnehmer saßen während der Studie auf einem Drehstuhl und nutzten ein Oculus Rift HMD. Zum Aktivieren von Regionen wurde der Oculus Controller verwendet.

In beiden Gruppen wurden *Cursor* und *Shape* als on-screen Indikator verwendet. Zusätzlich wurden Strahlen, bzw. Kantenspots als off-screen Indikatoren angezeigt, die auf Rols außerhalb des FoVs hinwiesen. Sowohl die on-screen- als auch die off-screen Anzeige wurden den Teilnehmern zuvor nicht erklärt und sollte intuitiv verstanden werden.

Es wurden die gleichen Fragebögen und Materialien wie in **Studie 7.3.3** genutzt.

7.3.4.2 Ergebnisse

VR Experience (Presence, Simulator-Sickness)

Die Scores der Presence-Items weisen für die beiden Methoden keine signifikanten Unterschiede auf.

Es gab kaum Simulator-Sickness. Die Unterschiede zwischen den Gruppen waren nicht signifikant.

User Experience

Beide Indikatoren waren hilfreich und verständlich, wobei der Strahl als verständlicher eingestuft wurde (Tabelle 7-11). Der Kantenspot wird als störender empfunden.

	störend	hilfreich	verständlich
Strahl	2.25 (1.28)	5.88 (1.13)	6.50 (0.75)
Kantenspot	3.25 (1.58)	5.63 (1.3)	6.0 (0.53)
p-Wert	0.04	0.24	0.04

Tabelle 7-11: Ergebnisse (Mittelwerte und Standardabweichungen) für die User-Experience-Items (7-Punkte-Likert-Skala). Die signifikant unterschiedlichen Werte (paarweiser t-Test, $p < 0.05$) sind fett markiert.

Effektivität

Zusätzlich zum Fragebogen wurden die Aktionen der Zuschauer aufgezeichnet. Mit Ausnahme von zwei Teilnehmern (je einer aus jeder Gruppe) verwendeten alle anderen jede Region, um zu den anderen Szenen zu wechseln.

7.3.4.3 Diskussion

Der Kantenspot wurde von den Teilnehmern als störend empfunden, was aber auch durch die gewählten Parameter wie Größe und Farbe verursacht worden sein kann. In unserem Experiment haben wir nicht untersucht, wie sich die Anzahl der gleichzeitig angezeigten off-screen Indikatoren auf das Zuschauererlebnis auswirkt. Es war jedoch erkennbar, dass off-screen Anzeigen besonders störend wirken, wenn mehrere off-screen Regionen gleichzeitig angezeigt werden. In zukünftigen Untersuchungen sollte nach weiteren Lösungen gesucht werden, die die Anzahl der anzuzeigenden Regionen berücksichtigen.

Die Methoden zum Anzeigen von on- und off-screen Regionen sind nicht nur für CVR relevant, sondern können auch in anderen VR- und AR-Anwendungen genutzt werden. Unsere Studien zur on- und off-screen Visualisierung haben vielversprechende Ergebnisse geliefert, die ermutigen, Forschungen in diese Richtung fortzusetzen.

7.3.5 Studie: Blickbasierte Interaktionstechniken

Die Studie ist Teil der Bachelorarbeit von Pascal Pothmann, (2019): Activation Methods in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und gemeinsam mit Pascal Pothmann, Heiko Drewes und Heinrich Hußmann diskutiert.

Teile dieser Studie sind vorab veröffentlicht worden:

- Rothe, S., Pothmann, P., Drewes, H., Hussmann, H., 2019d. Interaction Techniques for Cinematic Virtual Reality, in: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE

Diese Studie wurde durchgeführt, um einen Einblick darüber zu erhalten, wie sich blickbasierte Interaktionsmethoden in CVR auf Simulator-Sickness, Task Workload, Benutzererfahrung und Presence auswirken. Die Teilnehmer wendeten vier verschiedene **Interaktionsmethoden** an (Abbildung 7-78):

Blink (Blinzeln)

Bei der *blink*-Methode blinzelt der Betrachter zweimal, um einen Bereich zu aktivieren, ähnlich wie beim Doppelklick mit der Maus. Zwei aufeinanderfolgende Lidschläge werden als Signal erkannt, wenn das Zeitintervall zwischen ihnen weniger als 500 ms beträgt.

Click

Für die *click*-Methode wird ein in der Hand gehaltener Mini-Bluetooth-Controller genutzt. Der Betrachter schaut auf die RoI und drückt die Taste.

Nod (Nicken)

Bei der *nod*-Methode richtet der Betrachter den Blick auf das Ziel und nickt mit dem Kopf, während die Fixierung durch die Augen beibehalten wird (Mardanbegi u. a., 2012). Dazu muss der Kopf etwa 10 Grad nach unten und oben bewegt werden, damit das Nicken erkannt wird. Auch andere Kopfbewegungen wären möglich. Wir wählten das Nicken, da es in vielen Kulturen als Hinweis auf Übereinstimmung gilt.

Dwell

Um die Dwell-Zeit zu nutzen, richtet der Betrachter seinen Blick für ein bestimmtes Zeitintervall auf die RoI. Er erhält ein Feedback darüber, wie lange das Ziel bereits betrachtet wurde und wie lange es noch fixiert werden muss, bis die Interaktion ausgelöst wird. In unserem Fall war dieses Feedback ein sich füllender Kreis und der Vorgang begann mit einer Verzögerung von 300 ms, um Störungen zu vermeiden.

Basierend auf den Ergebnissen unserer Pilotstudie und den Experimenten von Kallioniemi et al. (Kallioniemi et al., 2017) wurde eine Dwell-Zeit von 2 s gewählt.

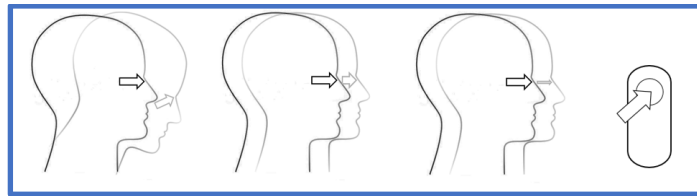


Abbildung 7-78: Die vier Selektionsmethoden von links nach rechts: (1) *nod*: Kopfnicken während die RoI mit den Augen fixiert wird. (2) *dwell*: Die RoI wird für 2 s mit den Augen fixiert. (3) *blink*: Die RoI wird mit den Augen fixiert und zweimal kurz hintereinander geblinzelt. (4) *click*: Die RoI wird mit den Augen fixiert und eine Controller-Taste gedrückt.

In allen Testfällen sind die Ränder der Interaktionsziele hervorgehoben (Farb- und Deckkraftänderung), wenn der Benutzer darauf schaute. Der Cursor war im Allgemeinen nicht sichtbar, außer in Verbindung mit dem Fortschrittsindikator der Dwell-Interaktion.

Die Parameter und Schwellwerte der Methoden wurden in einer informellen Pilotstudie ermittelt. Dabei fiel auf, dass diese sehr stark von einzelnen Anwendern und sogar von der Applikationsgestaltung abhängig sind. Wir sind uns bewusst, dass unsere Parameter nicht für jede Person geeignet sind. Trotzdem sind die Studienergebnisse wichtig, um Erfahrungen mit blick- und kopfbasierten Interaktionstechniken in CVR zu sammeln.

7.3.5.1 Material

Es wurden vier interaktive CVR-Applikationen produziert, die in Inhalt und Länge ähnlich sind. Die Videoszenen zeigen verschiedene Plätze in einem keltischen Museumsdorf mit Protagonisten als keltische Bewohner. Jede Applikation besteht aus zwei Videoszenen, in denen der Betrachter durch Interaktion mit einem Objekt (z. B. einer Tür) zur anderen Szene wechseln kann. Die Interaktionsbereiche wurden wie in **Studie 7.3.3** mit einem Rand versehen, um den Teilnehmern zu signalisieren, wo sie interagieren können (Abbildung 7-79).

Die Videoszenen haben eine Länge von ca. 2 Minuten und wurden erneut abgespielt, wenn der Benutzer die Szene vor dem Ende des Videos nicht gewechselt hat. Darüber hinaus gab es mehrere Objekte, die zusätzliche Informationen zur keltischen Geschichte boten.

Es wurden zwei **Aktionstypen** untersucht: (1) den Wechsel zur nächsten Szene und (2) ein Popup-Fenster mit zusätzlichen Informationen (Abbildung 7-80). Für jede Experience hatte die Eröffnungsszene zwei Interaktionsbereiche, einen für den Szenenwechsel und einen für zusätzliche Informationen. Die anderen Szenen hatten fünf Interaktionsbereiche und einen Szenenwechselbereich.

Für die Studie wurden ein Fove HMD mit integriertem Eyetracker und Kopfhörer von Bose mit aktiver Geräuschunterdrückung benutzt, für die *click*-Methode zusätzlich ein Mini-Bluetooth-Controller von auvisio.

Die virtuelle Umgebung wurde unter Verwendung von Unity 3D mit dem dazugehörigen FOVE-Plug-In entwickelt. Um die Sichtbarkeit der RoIs und Pointer anpassen zu können und damit flexibel für weitere Studien zu sein, wurden die folgenden Parameter so konzipiert, dass sie leicht geändert werden können:

- Pointer-Mount (Kopf oder Auge)
- Cursor-Darstellung (sichtbar / unsichtbar, Größe, Farbe, Deckkraft)
- Randdarstellung (sichtbar / unsichtbar, Farbe, Deckkraft)

Auf diese Weise konnten wir in unserer Pilotstudie nach den besten Parametern suchen und können das Projekt für weitere Forschungen nutzen.



Abbildung 7-79: Ein Rand zeigt die Tür als RoI, die die nächste Szene startet. Im Projekt ist der Rand subtiler, er wurde in dieser Abbildung zur Illustration stärker betont.

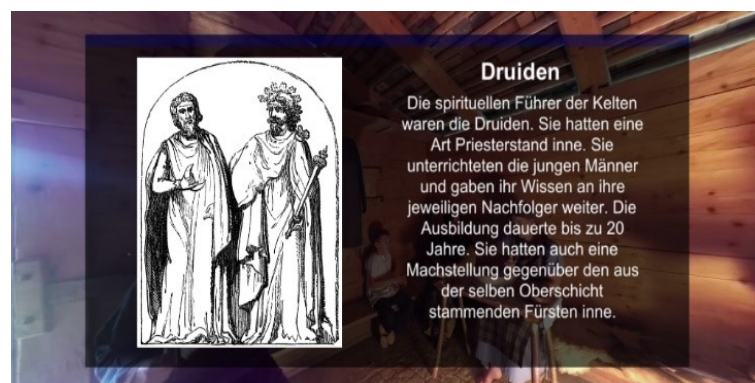


Abbildung 7-80: Auswahl eines Bereichs: Es können verschiedene Aktionen ausgelöst werden, z. B. Anzeigen zusätzlicher Informationen.

7.3.5.2 Teilnehmer und Prozedur

24 Personen (12 weiblich, 12 männlich, Alter: MW=22.8 Jahre) nahmen an der Studie teil. Zu Beginn beantwortete jeder Teilnehmer den demographischen Teil des Fragebogens und der Eyetracker wurde kalibriert. Daraufhin erhielten die Teilnehmer eine Einführung in die Interaktionsmethoden, und die Möglichkeit für fünf Minuten, eine Szene zu erkunden und mit jeder Methode zu interagieren. Diese Szene wurde in der Studie nicht wiederverwendet.

Es wurde ein within-subject Testdesign mit der Methode als unabhängige Variable verwendet. Jeder Teilnehmer schaute sich vier Filmerlebnisse an, jedes mit einer anderen Methode. Die Reihenfolgen der Methoden und Filmerlebnisse wurden ausbalanciert. Nach jedem Video wurden unter Verwendung etablierter Fragebögen Presence (IPQ) (Schubert u. a., 2002), Simulator-Sickness (SSQ) (Kennedy u. a., 1993) und Benutzererfahrung (UEQ) (Hinderks u. a., 2018) gemessen. Nachdem die Teilnehmer alle Videos angesehen hatten, wurden einige abschließende Fragen zum Vergleich der Methoden beantwortet, und es bestand die Möglichkeit, Gedanken zu den Methoden als Text aufzuschreiben.

7.3.5.3 Ergebnisse

Der Shapiro-Wilk-Test zeigte, dass nicht für alle Items Normalverteilung angenommen werden konnte. Daher wurde der Friedman-Test angewendet, um signifikante Unterschiede zwischen den Methoden zu finden. Ein α -Level von 0.05 wurde gewählt, um die Signifikanz zu bestimmen. In Fällen, in denen der Friedman-Test signifikante Unterschiede aufwies, wurden paarweise post-hoc Vorzeichen-Rang-Tests durchgeführt.

Presence

Es gab keine signifikanten Unterschiede zwischen den Methoden bezüglich der IPQ-Items. Die Fragen wurden auf einer 5-Punkte-Likert-Skala beantwortet und ergaben für die allgemeine Presence die folgenden Werte: $MW_{\text{blink}} = 3.42$, $SD_{\text{blink}} = 0.97$, $MW_{\text{click}} = 3.54$, $SD_{\text{click}} = 0.66$, $MW_{\text{nod}} = 3.58$, $SD_{\text{nod}} = 0.83$, $MW_{\text{dwell}} = 3.67$, $SD_{\text{dwell}} = 0.7$. Der Friedman-Test zeigte einen p-Wert von 0.86.

Simulator-Sickness

Generell gab es nur wenig Simulator-Sickness (0-0.75) und es konnten keine signifikanten Unterschiede zwischen den Methoden festgestellt werden. Die Frage nach dem allgemeinen Unbehagen ergab die folgenden Werte: $MW_{\text{blink}} = 0.42$, $SD_{\text{blink}} = 0.72$, $MW_{\text{click}} = 0.54$, $SD_{\text{click}} = 0.83$, $MW_{\text{nod}} = 0.46$, $SD_{\text{nod}} = 0.72$, $MW_{\text{dwell}} = 0.38$, $SD_{\text{dwell}} = 0.71$. Der Friedman-Test zeigte einen p-Wert von 0.9.

User Experience

Mehrere Items des UEQs, weisen signifikante Unterschiede auf. Bei den paarweisen post-hoc Tests ergaben sich signifikante Unterschiede für die Items, die in Tabelle 7-12 aufgeführt sind.

	best	worst	p-value
Comprehensibility	click	nod	0.001
Creative	blink	click	0.03
Easy To Learn	click	nod	0.0002
Fast	click	nod/dwell	0.001
Inventive	nod	click	0.002
Helpful	click	nod/dwell	<0.001
Easy To Use	click	nod	0.008
Novelty	blink	click	0.002
Enjoyable	click	nod	0.01
Stimulation	blink	dwell	0.005
Efficiency	click	dwell	<0.001

Tabelle 7-12: Items des UEQ mit signifikanten Unterschieden.

Vergleich

Auf einer 5-Punkte-Likert-Skala bewerteten die Teilnehmer für jede der Methoden verschiedene Aspekte der Experience: schnelle Interaktion, Ablenkung vom Video und komfortable Interaktion. Für alle diese Aspekte ergab die *click*-Methode die besten Ergebnisse (Tabelle 7-13).

	best	worst	p-value
fast interacting	click 4.46 (SD= 0.66)	dwell 2.5 (SD=0.93)	0.002
distraction from the video	click 1.54 (SD= 0.83)	nod 3.17 (SD=1.2)	<0.001
uncomfortable interacting	click 1.41 (SD= 0.58)	nod 2.66 (SD= 1.37)	0.001

Tabelle 7-13: Mittelwerte und Standardabweichungen der Vergleichsfragen.

Zusätzlich bewerteten die Teilnehmer jede Methode auf einer 9-Punkte-Likert-Skala von 1 (sehr schlecht) bis 9 (sehr gut). Die *click*-Methode ergab das beste Ergebnis ($MW_{\text{click}} = 7.46$, $SD_{\text{click}} = 1.72$), gefolgt von der *blink*-Methode ($MW_{\text{blink}} = 6.54$, $SD_{\text{blink}} = 2.43$). Die *nod*- und die *dwell*-Methode wurden schlechter eingestuft: ($MW_{\text{nod}} = 3.75$, $SD_{\text{nod}} = 2.52$, $MW_{\text{dwell}} = 3.54$, $SD_{\text{dwell}} = 2.12$).

7.3.5.4 Qualitative Ergebnisse

Im letzten Teil des Fragebogens konnten die Teilnehmer Kommentare zu den einzelnen Methoden notieren.

Die meisten Teilnehmer hatten Probleme mit der *nod*-Methode. Auch wenn sie „die Idee gut“ (P1) fanden, „funktionierte sie nicht immer genau“ (P1, P3, P6, P10, P20). Einige Teilnehmer fühlten sich nickend seltsam (P16, P18, P23). Außerdem ist das Nicken mit einem HMD (P7) schwieriger und störend (P3, P8). Die Bewegung des Kopfes kann „Schwindel verursachen“ (P12). Nur für einen Teilnehmer hat diese Methode „gut und schnell funktioniert“ (P21).

Für manche Teilnehmer hat die *dwell*-Methode „am besten funktioniert“ (P3), für andere „hat es zu lange gedauert“ (P1, P4, P6, P11, P12, P16, P21, P24). P19 erwähnte, dass „der Füllkreis die Experience stört“.

Die *blink*-Methode arbeitete für die meisten Teilnehmer „schnell“ (P1) und „effizient“ (P10, P11, P18). Für andere Teilnehmer war es „anstrengend“ (P20) und „ungewöhnlich“ (P12).

Die *click*-Methode war „einfach zu bedienen“ (P1, P9, P10, P11, P16, P24) und „schnell“ (P9, P16, P21). Bei einigen Teilnehmern hat der Controller jedoch die Presence verringert (P1, P10, P23) und „es ist störend, etwas in der Hand zu haben“ (P18, P20, P21, P23).

7.3.5.5 Diskussion

Die Ergebnisse dieser Studie werden in Abschnitt 4.2.2.2. diskutiert und dort für das Spaceline-Konzept die dazugehörigen Guidelines genutzt.

7.3.6 Studie: Kameraausrichtungen beim Spacewechsel

Die Studie ist Teil der Bachelorarbeit von Luca Speeter (2019): Camera Positions in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Diese Nutzerstudie setzte sich mit der 180°-Regel (Achsensprung) auseinander und untersuchte, was diese Regel für CVR bedeutet (siehe Abschnitt 4.2.3.4).

Baseline

Für die Baseline wurde die Kamera am Ende eines Tisches platziert, an dem zwei Personen miteinander sprachen (Abbildung 7-81, B). Der Betrachter konnte beide Charaktere gleichzeitig sehen.

Für die beiden anderen Bedingungen wurde die Kamera neben die Charaktere platziert und während der Schnitte zwischen den beiden Charakteren gewechselt (Abbildung 7-81, A1, A2). Zusätzlich zur Kameraposition wurde die Richtung für beide Methoden auf unterschiedliche Weise geändert. Dabei muss berücksichtigt werden, dass die Blickrichtung nur in horizontaler Richtung zugewiesen werden kann. Alle vertikalen Änderungen würden Zenit und Nadir der virtuellen Welt verschieben und nicht mit der Körperhaltung des Betrachters übereinstimmen.

personView

Bei der *personView*-Methode nimmt der Zuschauer die Position einer Person in der virtuellen Welt ein und die Blickrichtung wird an der virtuellen Welt ausgerichtet. In unserer Studie war der Tisch die Referenz für die Richtung. Der Teilnehmer saß vor dem Schnitt auf einer Seite des Tisches und die Tischkante war vor ihm (Abbildung 7-81, links, A1). Nach dem Schnitt saß er an der gegenüberliegenden Seite (Abbildung 7-81, links, A2).

toRoI

Bei der *toRoI*-Methode ist die Blickrichtung nach dem Schnitt zur RoI ausgerichtet, in unserem Experiment zur aktuell sprechenden Person, unabhängig davon, wohin der Betrachter zuvor geschaut hat. Die Kameraposition wechselte zwischen zwei gegenüberliegenden Stellen des Tisches (Abbildung 7-81, rechts, A1 und A2). Beide Positionen befanden sich auf der gleichen Seite der sprechenden Charaktere. Diese Methode entspricht den viewpoint-orientierten Schnitten von Pavel et al. (Pavel et al., 2017). Wir haben einen anderen Namen verwendet, um mit unseren anderen Methoden konsistent zu sein.

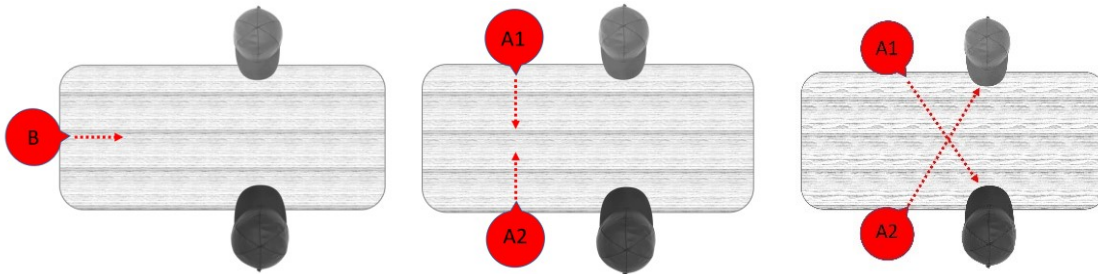


Abbildung 7-81: Die beiden Figuren sowie der Betrachter (Kamera) sitzen an einem Tisch. links: Kameraposition der Baseline (B). Mitte: Kamerapositionen der *personView*-Methode (A1, A2). Rechts: Kamerapositionen der *toRoI*-Methode (A1, A2).

7.3.6.1 Material

Für die Tests wurden drei kurze, sphärische Videos mit einer Länge von ca. 3 min erstellt. Diese waren in Stil und Inhalt sehr ähnlich: Zwei Personen sitzen an einem Tisch und unterhalten sich über unterschiedliche Dinge. Der Raum und die Charaktere sind in allen Videos gleich, nur der Inhalt der Unterhaltung ändert sich. Auf jedes der drei Videos wurden die drei Methoden angewendet, was zu neun verschiedenen Anwendungen führte. Die Methoden wurden in Unity 3D implementiert und die Videos mit einer Samsung Gear VR gezeigt (Samsung, 2018).

7.3.6.2 Teilnehmer und Prozedur

21 Personen (4 weiblich, 17 männlich, Alter: MW= 30.9, SD = 11.7) nahmen an der Studie teil. Es wurde ein within-subject Testdesign verwendet, und jeder Teilnehmer sah alle drei Testfälle jeweils für ein anderes Video. Die Reihenfolge der Varianten und Videos wurde durch ein Graeco-Latin-Quadrat ausgeglichen. Bevor die Tests begannen, schauten sich die Teilnehmer ein kurzes Video an, um sich an das Medium zu gewöhnen. Dieses Video hatte eine Länge von 2:50 min und zeigte den gleichen Raum wie die Testvideos, jedoch ohne Schnitte. Die Teilnehmer wurden gebeten, sich umzusehen, um den Novelty-Effekt zu minimieren und eine eingehendere Betrachtung des Raums im ersten Testfall zu vermeiden. Die Dauer dieser Einführung wurde von den Teilnehmern bestimmt und betrug durchschnittlich 1 min.

Zu Beginn beantwortete jeder Teilnehmer Fragen zu Alter, Geschlecht und VR-Erfahrung. Nach jedem Film wurden Fragebögen ausgefüllt, um die Presence (Schubert u. a., 2002) und die Simulator-Sickness (SSQ) zu messen (Kennedy u. a., 1993). Zusätzlich wurden spezielle Fragen zur Usability für die drei Methoden gestellt. Am Ende beantworteten die Teilnehmer einige abschließende Fragen zum Vergleich der Methoden.

7.3.6.3 Ergebnisse

Der Shapiro-Wilk-Test zeigte, dass nicht für alle Items Normalverteilung angenommen werden konnten. Daher wurde der Friedman-Test angewendet, um signifikante Unterschiede zwischen den Bedingungen zu finden (α -Level von 0.05).

Presence

Für die Presence-Items konnten keine signifikanten Unterschiede zwischen den Bedingungen festgestellt werden. Die Frage nach der allgemeinen Presence ergab folgende Werte: $p=0.44$ (Friedman), $MW_{personView}=6.62$, $SD_{personView}=1.5$, $MW_{base}=5.33$, $SD_{base}=1.46$, $MW_{toRoI}=5.14$, $SD_{toRoI}=1.56$.

Simulator-Sickness

Es wurde allgemein nur wenig Simulator-Sickness (0-0.89) festgestellt. Das Item über allgemeine Beschwerden zeigte höhere Scores für die *personView*-Methode: $p=0.046$ (Friedman), $MW_{personView}=0.86$, $SD_{personView}=0.85$, $MW_{base}=0.38$, $SD_{base}=0.59$, $MW_{toRol}=0.29$, $SD_{toRol}=0.56$ (paarweise post-hoc Tests: $p_{base/personView}=0.049$, $p_{personView/toRol}=0.01$). Die anderen Fragen zeigten keine signifikanten Unterschiede.

Experience

Die Fragen wurden auf einer 5-Punkte-Likert-Skala beantwortet. Bei vier der sieben Fragen konnten signifikante Unterschiede festgestellt werden: (E2) „Ich musste meinen Kopf bewegen, um zu folgen“, (E4) „Die Schnitte haben mich vom Inhalt abgelenkt“, (E5) „Die Schnitte haben geholfen, dem Inhalt zu folgen“, (E7) „Die Schnitte haben natürlich funktioniert“. Abbildung 7-82 und Tabelle 7-14 zeigen die Mittelwerte der Scores und die zugehörigen p-Werte der Friedman-Tests, sowie die Ergebnisse der paarweisen post-hoc Tests. Da sich die Fragen E4-7 auf Schnitte beziehen, gibt es keine Antwort für die Baseline

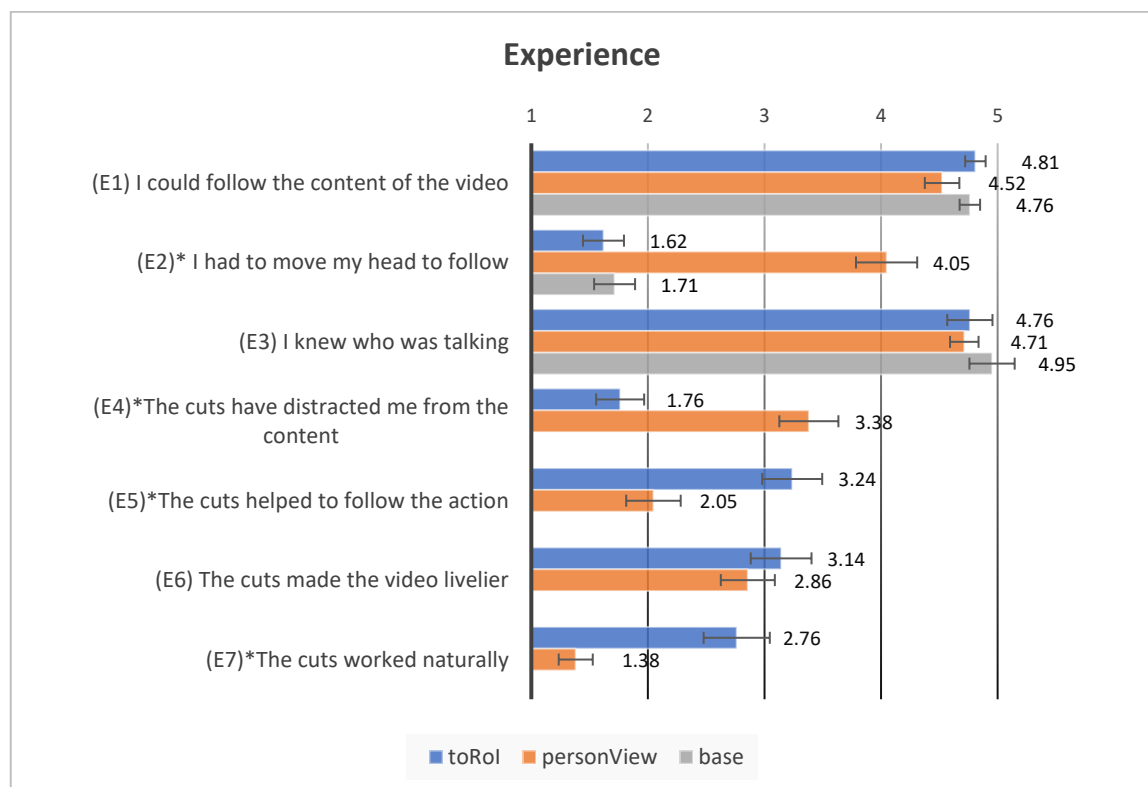


Abbildung 7-82: Mittelwerte und Standardfehler der Experience-Items für jede Bedingung (Likert-Skala von 1 bis 5).

	base	personView	toRol	pFriedmann	p _{base/personView}	p _{base/toRol}	p _{personView/toRol}
E2	1.71(0.78)	4.05(1.2)	1.61(0.8)	<0.001	<0.001	0.66	<0.001
E4	-	3.38(1.2)	1.76(0.9)	<0.001	-	-	<0.001
E5	-	2.05(1.1)	3.24(1.2)	0.001	-	-	0.002
E7	-	1.38(0.7)	2.76(1.3)	<0.001	-	-	<0.001

Tabelle 7-14: Mittelwerte und Standardabweichungen (in Klammern) der Experience-Items, die signifikante Unterschiede aufwiesen. Die letzten Spalten zeigen den p-Wert des Friedman-Tests und die p-Werte der paarweisen post-hoc Tests.

Vergleich

Die meisten Teilnehmer fanden das Baseline-Video am natürlichsten und niemand war davon verwirrt (Abbildung 7-83). Im Gegensatz dazu fühlte sich das *personView*-Video für keinen der Teilnehmer natürlich an, 90 % waren davon verwirrt. Bei der Frage, welche Variante für ein längeres Video am besten geeignet ist, wählte etwa die Hälfte der Teilnehmer das Baseline-Video (48 %) und der andere Teil (52 %) die *toRoI*-Methode. Das bessere Ergebnis für die *toRoI*-Methode im Vergleich zur Baseline kann durch die Tatsache verursacht sein, dass ein längerer Film ohne Schnitte langweilig wird. Dies scheint auch für CVR-Videos zuzutreffen, bei denen die Teilnehmer den Bildausschnitt durch Bewegen des Kopfes selbst ändern.

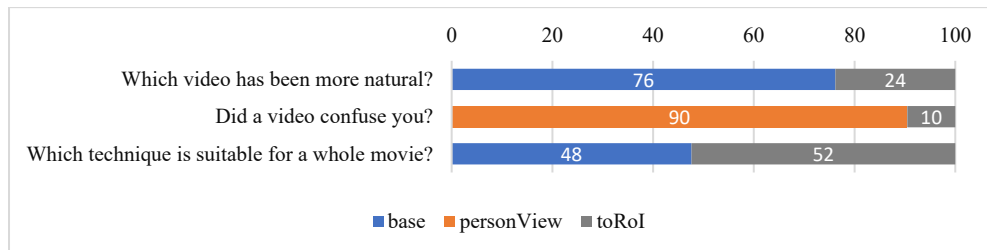


Abbildung 7-83: Results of the comparing part (in percent).

7.3.6.4 Diskussion

Die Ergebnisse dieser Studie werden in Abschnitt 4.2.3.4. diskutiert und dort für das Spaceline-Konzept und die dazugehörigen Guidelines genutzt.

7.4 Nutzerstudien: Nutzerkomfort

Um die Nutzung einer CVR-Applikation so angenehm wie möglich zu machen führten wir mehrere Studien durch und betrachteten dabei verschiedene Aspekte: Die Studien in 7.4.1 setzten sich mit den Möglichkeiten auseinander, CVR gemeinsam anzuschauen. Um eine CVR-Applikation auch ohne ständiges Umherdrehen anschauen zu können, werden in den **Studien 7.4.2 bis 7.4.4** verschiedene Ansätze miteinander verglichen. Die **Studien 7.4.5 bis 7.4.8** analysieren Methoden, Untertitel und Gebärdendolmetscher in CVR zu integrieren.

7.4.1 Studie: Social Viewing in CVR

Die Studie ist Teil der Bachelorarbeiten von

- Alexander Schmidt (2018): Collaborator Awareness in Cinematic Virtual Reality
- Viet Duc Ngo (2018): The Influence of Viewport Sharing on Social Awareness in CVR
- Vo Duc Dung (2018): Methods of Communication for Social Viewing in Cinematic VR
- Atef Azabi (2018): Influence of Voice Chat for Social Viewing in Cinematic VR

Ich habe diese Arbeiten betreut und gemeinsam mit den Studenten die Ideen entwickelt, die sie implementierten und testeten. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir von den Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

In dieser Studie wurden vier Ansätze untersucht, um das gemeinsame Schauen von CVR-Filmen zu ermöglichen: Voice-Chat, Senden von Emotionszuständen, Anzeige von Sichtfeldern und Video-Chat. Für jede der Komponenten wurden zwei Methoden miteinander verglichen. Ziel der Studie war es, herauszufinden, welche Methoden für Social Viewing geeignet sind, und mehr über die Vor- und Nachteile der einzelnen Methoden zu erfahren.

Voice-Chat

Es wurden zwei Voice-Chat Methoden getestet. Die erste Methode verwendete räumlichen Sound (*spatial*-Methode), während die zweite normalen Stereoklang nutzte (*nonspatial*-Methode), der nicht von der Blickrichtung des Betrachters abhing. Der räumliche Klang kam aus der Richtung, in die der sprechende Teilnehmer schaute, und basierte auf Headtracking-Daten.

Video-Chat

Die beiden Methoden für den Video-Chat unterscheiden sich in der Position des Chat-Fensters. In der *front*-Methode ist der Chat-Screen mit dem Co-Watcher mittig am unteren Displayrand fixiert, auch wenn der Kopf bewegt wird (Abbildung 7-84, links). Der Co-Watcher ist von vorn zu sehen und die Situation ähnelt der Situation, in der sich zwei Personen gegenüber sitzen. Das Chat-Fenster bleibt jedoch immer im FoV. Bei der zweiten Methode, der *side*-Methode, ist das Video-Chat-Fenster und die Kamera neben dem Zuschauer platziert (Abbildung 7-84, rechts). In diesem Fall hängt es von der Blickrichtung ab, ob das Chat-Fenster zu sehen ist. Dieser Fall entspricht der Situation des Nebeneinandersitzens beim gemeinsamen Filmschauens. Auch wenn man für eine reale Anwendung einen Video-Chat mit einem Voice-Chat kombinieren würde, wurde die Methode ohne Sprache getestet, da der Einfluss jeder Komponente einzeln untersucht werden sollte.



Abbildung 7-84: Zwei Methoden zum Anzeigen des Co-Watchers, links: *front*-Methode, rechts: *side*-Methode.

Viewport-Indikation

Um das Bewusstsein für den Viewport des anderen zu untersuchen, haben wir zwei Methoden ausgewählt: *bar*-Methode und die *PiP*-Methode (Picture-in-Picture). Bei der *PiP*-Methode war das Sichtfeld der anderen Person in einem kleinen Videofenster zu sehen, dessen Position sich beim Umherschauen nicht veränderte (screen-referenziert) (Abbildung 7-85, links). Die *bar*-Methode ist von einem Segelflugwarnsystem inspiriert. Am unteren Rand des Displays und auf der rechten Seite befinden sich Balken entlang einer Linie (Abbildung 7-85, rechts). Einer der Balken ist rot und zeigt an, ob das FoV der anderen Person links, rechts, oberhalb oder unterhalb des eigenen FoVs liegt.

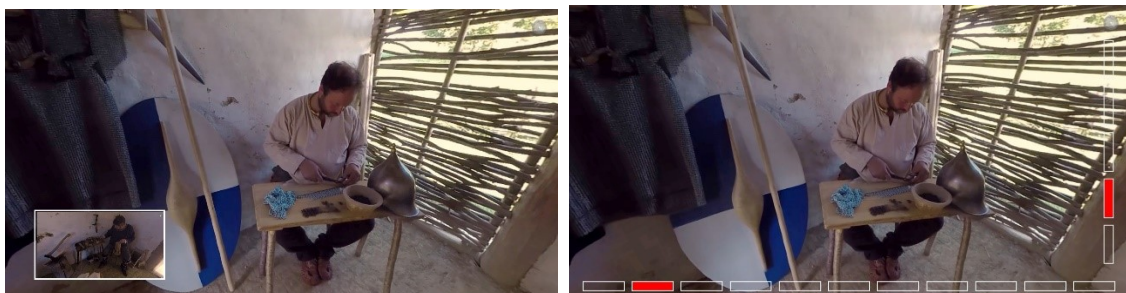


Abbildung 7-85: Zwei Techniken, zum Anzeigen des Viewports, links: *PiP*-Methode, rechts: *bar*-Methode.

Senden von Emotionszuständen

Für die Übermittlung von Emotionen wurden zwei visuelle Methoden implementiert: das Senden von *Smileys* und das Senden von *Fotos* mit verschiedenen Gesichtsausdrücken (Abbildung 7-86). Für beide Methoden standen je vier Bilder zur Verfügung: zwei positive und zwei negative. Zum Senden der Bilder wurde ein Mini-Controller verwendet. Die verwendeten Fotos zeigten keine bekannten Personen und

könnten in Zukunft durch Fotos ersetzt werden, die im Voraus aufgenommen werden und den Partner selbst zeigen. Die Position des Bildes war immer die gleiche, am Rand des Displays. Diese Methode kann durch Positionieren des Bildes auf die Rol erweitert werden. Dies würde jedoch zusätzliche Lenkungsmethoden erfordern, falls die Rol nicht im FoV enthalten ist.



Abbildung 7-86: Smileys oder Fotos zur Information über die emotionalen Zustände des Co-Watchers.

7.4.1.1 Material

Um erste Erfahrungen zu sammeln, konzentrierten wir uns in unseren Studien auf zwei Personen in symmetrischen VR-Umgebungen (beide mit HMDs), die gemeinsam an einer CVR-Applikation teilhaben. Alle Methoden wurden so implementiert, dass sie für *remote* und *co-located* Umgebungen verwendet werden können. Für alle Tests nutzten die Teilnehmer eine Oculus Rift mit Kopfhörern.

Mit Ausnahme des Voice-Chat-Tests wurden alle Tests in einem großen Raum durchgeführt und eine remote Umgebung simuliert: Beide Teilnehmer waren nur über das Netzwerk miteinander verbunden. Wir wählten diese Einraum-Konstellation, um beide Teilnehmer gleichzeitig beobachten zu können. Dies war nur für die visuellen Methoden möglich, da das HMD die visuelle reale Umgebung blockiert. Die Stimmen waren jedoch noch wahrnehmbar, auch wenn die Teilnehmer Kopfhörer verwendeten. Für den Voice-Chat-Test fand die Studie deshalb in zwei benachbarten Räumen statt, so dass sich die Teilnehmer während der Studie in verschiedenen Räumen aufhielten.

7.4.1.2 Teilnehmer und Prozedur

Die Studie bestand aus vier Komponenten mit einer unterschiedlichen Anzahl an Teilnehmern. Für jede Komponente wurden zwei verschiedene Methoden miteinander verglichen (Tabelle 7-15).

Komponente	Teilnehmer	Methode 1	Methode 2	Ansatz für
Voice-Chat	23 Teilnehmer 4 weiblich, 19 männlich Alter MW=27.26	nonspatial	spatial	Kommunikation Soziale Presence Barrierefreiheit Viewport Awareness
Video-Chat	22 Teilnehmer 7 weiblich, 15 männlich Alter MW =30.7	front	side	Kommunikation Soziale Presence Barrierefreiheit
Viewport-Indikation	21 Teilnehmer 8 weiblich, 13 männlich Alter MW =23.9	PiP	bar	Viewport-Awareness Soziale Presence
Emotionen	20 Teilnehmer 6 weiblich, 14 männlich Alter MW =25.6	smiley	photo	Kommunikation Soziale Presence Barrierefreiheit Viewport Awareness Interaktion

Tabelle 7-15: Für jede Komponente wurden zwei Methoden implementiert und miteinander verglichen. Die letzte Spalte gibt an, für welche Herausforderungen die Methoden hilfreich sind.

Jeder Teilnehmer sah zwei Filme mit unterschiedlichen Methoden. Die Filme und Methoden wurden in Reihenfolge und Zuordnung ausbalanciert. Alle Filme waren Naturdokumentationen in 4K-Auflösung, ähnlich in Stil und Tempo und hatten eine ungefähre Länge von 8 Minuten. Die unabhängige Variable war die Methode, die abhängigen Variablen Presence, Simulator-Sickness, Usability und Zusammengehörigkeitsgefühl.

Um Aussagen über die soziale Awareness zu erhalten, verwendeten wir die folgenden Fragen aus dem ABC-Fragebogen (IJsselsteijn u. a., 2009):

- (S1) Ich fühle mich aufgrund der Kontakte als Teil einer Gruppe.
- (S2) Ich weiß, was der andere im Moment empfindet.
- (S3) Aufgrund der Kontakte kann ich mich mit den anderen identifizieren.

Es wurde eine 7-Punkte-Likert-Skala zur Bewertung der Antworten verwendet. Am Ende der Studie beantworteten die Teilnehmer Fragen zum Vergleich der Methoden:

- (C1) Welche Methode finden Sie komfortabler (Usability)?
- (C2) Würden Sie die Methode länger nutzen (Usability)?
- (C3) Mit welcher Methode fühlen Sie sich mehr mit dem anderen verbunden (Zusammengehörigkeit)?

In einem zweiten Teil verglichen wir die einzelnen Komponenten miteinander (between-subject), um herauszufinden, welche von ihnen den größten Beitrag zum sozialen Seherlebnis leisten und welche eine untergeordnete Rolle spielen und möglicherweise weggelassen werden können.

7.4.1.3 Ergebnisse: Vergleich der Methoden für jede Komponente

Um die beiden Methoden jeder Komponente miteinander zu vergleichen, führten wir für jedes Likert-Item ein t-Test durchgeführt ($\alpha = 5\%$), bei dem keine signifikanten Unterschiede in Bezug auf Presence, Simulator-Sickness festgestellt wurden. Für die statistische Auswertung der Vergleichsfragen wurde der exakte Fisher-Test verwendet. Aus den Ergebnissen der Vergleichsfragen (C1-C3) und der qualitativen Analyse der textlichen Anmerkungen ermittelten wir Präferenzen, Vor- und Nachteile, die im Folgenden detailliert dargestellt werden:

Voice-Chat: Beide Methoden wurden von den Teilnehmern akzeptiert. Die meisten würden die Methoden auch für längere Videos nutzen (78.2 %, 87 %). In allen Fragen erhielt die räumliche Methode größere Zustimmung (Tabelle 7-16).

	komfortabler	für eine längere Zeit nutzbar	mehr Verbundenheit mit den anderen
nonspatial	30.4 %	78.2 %	26.1 %
spatial	69.6 %	87 %	73.9 %

Tabelle 7-16: Vergleich der beiden Methoden. Die räumliche Voice-Chat Methode wurde von den meisten Teilnehmern bevorzugt.

Die meisten Teilnehmer bevorzugten Voice-Chat mit räumlichem Sound. P14 verglich die beiden Methoden folgendermaßen: „Bei der ersten Methode (*nonspatial*) drehte sich der Großteil des Gesprächs um die Blickrichtung. Mit der zweiten Methode (*spatial*) konnte ich das allein an der Richtung erkennen, aus der seine Stimme kam, und man konnte sofort über die Bilder sprechen, anstatt sie erst finden zu müssen“. P22 bevorzugte die räumliche Methode: „Sie hat sich durch die unterschiedlichen Positionen des Tons stärker in das Video integriert gefühlt. Dies machte die Experience interessanter“. Einige Teilnehmer nannten Vor- und Nachteile der räumlichen Methode: „Die räumliche Methode war hilfreich, um die Sicht des anderen zu finden. Es war jedoch auch etwas ablenkender“(P17). „Die Stereomethode lenkt weniger ab. Mit der räumlichen Methode willst du auch schauen, wo der andere hinschaut und suchst so immer die richtige Blickrichtung“. Für ein besseres Verständnis von CVR als soziale Erfahrung sind die folgenden Gründe relevant.

Die meisten Teilnehmer bevorzugten die *spatial*-Methode:

- „Die Methode hat mir geholfen, die Ansicht des anderen zu finden“ (P8, P11, P14, P17, P18, P19, P21)
- „Die Erfahrung war räumlicher und realer.“ (P10, P21, P23)
- „Es war ein Gefühl, im selben Raum zu sein.“ (P11, P22)
- „Es fühlte sich eher so an, als würde man eine Umgebung gemeinsam erkunden.“ (P11)
- „Es gab mehr Interaktion.“ (P13)
- „Es war näher an der realen Erfahrung. (P14).

Einige Teilnehmer bevorzugten die *nonspatial*-Methode:

- „Es ist vertrauter.“ (P15, P16)
- „weniger verwirrend“ (P16)
- „weniger ablenkend“ (P16, P17)
- „Man kann gut hören.“ (P5, P6).

Video-Chat: Es trat ein signifikanter Unterschied für Item (S2) auf, bei dem der Score für die *side*-Methode ($MW_{side} = 4.41$, $SD_{side} = 1.87$) signifikant höher war als für die *front*-Methode ($MW_{front} = 3.27$, $SD_{front} = 1.78$, $p_{side/front} = 0.04$).

In diesem Teil der Studie unterschieden sich die Bewertungen für Benutzerfreundlichkeit und Zusammengehörigkeit für beide Methoden nicht signifikant (Tabelle 7-17).

	komfortabler	für eine längere Zeit nutzbar	mehr Verbundenheit mit den anderen
front	45.5 %	68.2 %	50.0 %
side	45.5 %	68.2 %	45.5 %

Tabelle 7-17: Vergleich der beiden Methoden. Es gab zwischen den beiden Video-Chat Methoden keinen signifikanten Unterschied. Einige Teilnehmer haben sich nicht für eine der Methoden entschieden.

Bei der Analyse der qualitativen Daten konnten wir wichtige Erkenntnisse für beide Methoden gewinnen. P18 bemerkte: „Es kommt auf den Inhalt an. Um einen Kinofilm anzusehen, finde ich die *side*-Methode besser.“

Die *front*-Methode war in folgenden Punkten überlegen:

- „ist bequemer“ (P2, P6)
- „ist besser für die Kommunikation“ (P11, P17, P20, P21)

Gründe für den Vorzug der Seitenansicht waren:

- „ähnlich wie Kino / TV“ (P1)
- „man kann mehr vom Partner sehen, wenn er nach vorne schaut“ (P1)
- „Man fühlt sich angesprochen, wenn er sich zu einem wendet“ (P1)
- „stört weniger, ist realistischer“ (P4, P7, P10, P11, P14, P19)

Viewport-Indikation: Auch im Viewport-Awareness Teil konnten wir keine signifikanten Unterschiede in Bezug auf die Benutzerfreundlichkeit zwischen den beiden Methoden feststellen. Bezüglich der Zusammengehörigkeit wurde die *PiP*-Methode jedoch von mehr Teilnehmern bevorzugt (Tabelle 7-18).

	komfortabler	für eine längere Zeit nutzbar	mehr Verbundenheit mit den anderen
PiP	47.6 %	71.4 %	52.4 %
bar	42.4 %	71.4 %	28.6 %

Tabelle 7-18: Vergleich der beiden Methoden. Bezüglich der Zusammengehörigkeit bevorzugten mehr Teilnehmer die *PiP*-Methode. Einige Teilnehmer haben sich für keine der Methoden entschieden.

Aus den textlichen Anmerkungen wurden Vorteile beider Methoden erkennbar:

Vorteile der *PiP*-Methode:

- „leichter zu verstehen und schneller“ (P3, P4, P18)
- „beide Ansichten können gleichzeitig gesehen werden“ (P8, P12, P16, P18)
- „ich sehe, was der andere sieht“ (P19)
- „persönlicher“ (P10)
- „verbundener“ (P15)
- „muss den Kopf nicht drehen“ (P10)

Vorteile der *bar*-Methode:

- „verdeckt weniger“ (P1)
- „weniger aufdringlich / diskret“ (P5, P6, P7, P11, P20, P21)
- „leichter zu verstehen“ (P5, P16, P20)
- „bessere Orientierung“ (P14)

Senden von Emotionszuständen: Die meisten Teilnehmer empfanden die *smiley*-Methode komfortabler (75 %). Die Mehrheit würde die *smiley*-Methode auch über längere Zeit nutzen (85 %). Das Zusammengehörigkeitsgefühl unterschied sich jedoch nur geringfügig. (Tabelle 7-19)

	komfortabler	für eine längere Zeit nutzbar	mehr Verbundenheit mit den anderen
smiley	75 %	85 %	55 %
photo	20 %	55 %	40 %

Tabelle 7-19: Vergleich der beiden Methoden. Die meisten Teilnehmer bevorzugten die *smiley*-Methode. Einige Teilnehmer haben sich nicht für eine der Methoden entschieden.

Für einige Teilnehmer schien die *photo*-Methode vertrauter, für andere waren es die Smileys. P20 wies darauf hin, dass Bilder von Gesichtern Erwartungen für die Kommunikation wecken könnten. Mehrere Teilnehmer gaben an, dass ein Foto realistischer ist (P3, P13, P16).

Vorteile von Smileys:

- „schneller und leichter zu erkennen“ (P2, P3, P5, P6, P8, P11, P16)
- „vertraut“ (P2, P18)
- „anonym bei Unbekannten“ (P7, P12)
- „weniger ablenkend“ (P13, P18)

Vorteile von Fotos:

- „authentischer“ (P4)
- „leichter zu verstehen“ (P9)
- „vertraut“ (P19)

7.4.1.4 Ergebnisse: Vergleich der Komponenten

Im zweiten Teil der Studie wurde untersucht, welche Komponenten für das Zusammengehörigkeitsgefühl wichtig sind. Dazu wurden die bevorzugten Methoden jeder Komponente aus dem ersten Teil miteinander verglichen:

- Voice-Chat: räumliche Methode (die Stimme des Co-Watchers kommt aus der Blickrichtung)
- Video-Chat: *front-Methode* (das Video-Chat-Fenster ist vor dem Benutzer mit dem Display verbunden)
- Viewport-Indikation: Anzeigemethode (ein PiP-Fenster zeigt den FoV des anderen Mitbeobachters)
- Senden von Emotionszuständen: *smiley*-Methode (Smileys zeigen die Emotionszustände an)

Ein Shapiro-Wilk-Test zeigte, dass nicht für alle Komponenten Normalverteilung vorausgesetzt werden konnten. Daher wurde der Friedman-Test verwendet, um signifikante Unterschiede zwischen den Komponenten zu finden. Es gab keine signifikanten Unterschiede in Bezug auf Presence und Simulator-Sickness. Bei den paarweisen post-hoc Tests fanden wir aber signifikante Unterschiede im Hinblick auf den Aspekt der sozialen Awareness (Fragen S1-S3). In Abbildung 7-87 sind die Mittelwerte und p-Werte zusammengefasst und dargestellt. Für signifikante Unterschiede ($p < 0.05$) werden die p-Werte in Schwarz hinzugefügt, für schwache signifikante Unterschiede ($0.05 \leq p < 0.1$) in Grau.

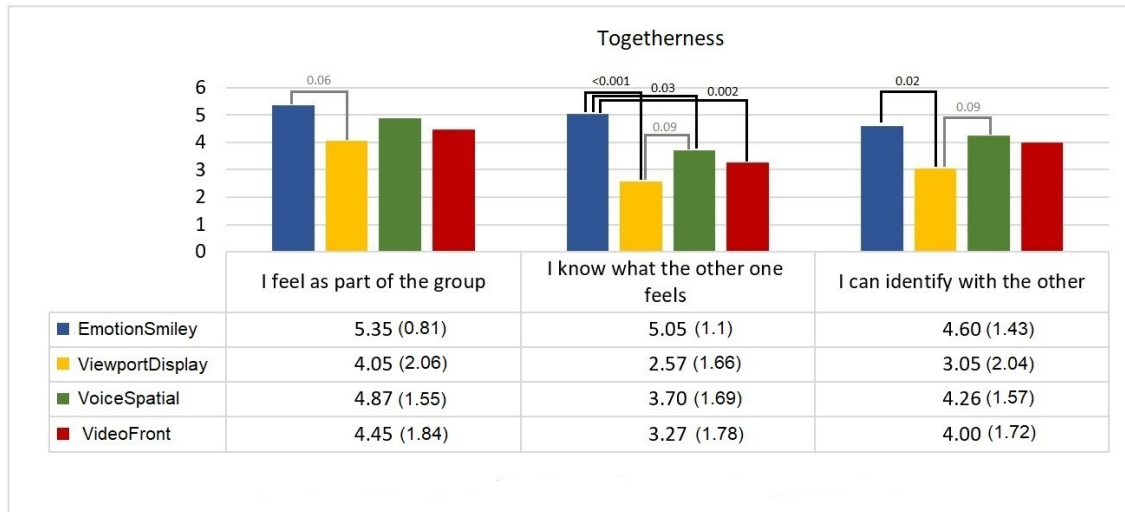


Abbildung 7-87: Mittelwerte und Standardabweichungen in Klammern für die Zusammengehörigkeit. Für signifikante Unterschiede ($p < 0,05$) sind die p-Werte in Schwarz angegeben, für schwache signifikante Unterschiede ($0,05 \leq p < 0,1$) in Grau.

Wie Abbildung 7-87 zeigt, war die *smiley*-Methode für die soziale Awareness am wichtigsten, gefolgt vom räumlichen Voice-Chat und dem Video-Chat. Die Teilnehmer unserer Studie erachteten es als weniger bedeutsam, das FoV des Mitbeobachters zu kennen.

7.4.1.5 Einschränkungen und Ausblick

Einige der Teilnehmer hatten noch nie einen Film mit einem HMD gesehen. Selbst für die anderen kann sich das Sehverhalten im Laufe der Zeit ändern, wenn öfter CVR-Videos angeschaut wurden.

In unserer Studie wurde nur ein Filmgenre (Naturdokumentationen) verwendet, und der Inhalt hat möglicherweise die Ergebnisse beeinflusst. Andere Arten von Filmen können mehr oder weniger Austausch und Interaktion erfordern. Weitere Untersuchungen sind notwendig, wie Stil und Tempo eines Films die sozialen Betrachtungsmethoden beeinflussen.

Wir untersuchten nur den Fall, in dem zwei Personen die Erfahrung miteinander teilen. Bei mehr Teilnehmern ist mehr Aufwand notwendig, um eine Überfüllung des Displays zu vermeiden. Zuordnungsmechanismen sind erforderlich, um zu identifizieren, welche Informationen zu welcher Person gehören. Ein Ansatz wäre ein Rollenkonzept.

Für die Eingabe der Smileys und Fotos verwendeten wir einen 4-Tasten-Controller und waren daher auf vier Smileys beschränkt. Wenn Gesten als Eingaben verwendet werden, sind mehr Smileys möglich. Die Smileys wurden unabhängig von der Blickrichtung des Absenders am Rand des Displays platziert. Diese Methode kann verbessert werden, indem der Smiley mit der Information verknüpft wird, die die Emotion auslöst. Befindet sich die Blickrichtung des sendenden Partners im FoV des Empfängers, könnte der Smiley auf dem betrachteten Objekt platziert werden. Befindet sich dieses Objekt außerhalb des Sichtfelds, kann der Smiley an der Displaykante platziert werden, die sich näher am Objekt befindet.

Da in unserer Studie alle Komponenten separat untersucht wurden, sollte in einem nächsten Schritt herausgefunden werden, wie sich die Komponenten am besten kombinieren lassen. Eines der Ergebnisse unserer Studie ist, dass Voice-Chat für die soziale Awareness sehr hilfreich ist und durch eine visuelle Kommunikation ergänzt werden kann.

Wir haben die soziale Beziehung zwischen den Teilnehmern nicht berücksichtigt. Einige von ihnen kannten sich vorher, andere nicht. Zwischen diesen beiden Optionen zu unterscheiden und die Fälle getrennt zu untersuchen, kann andere, neue Erkenntnisse hervorbringen.

7.4.2 Studie: Immobile Techniken

Die Studie ist Teil der Masterarbeit von Lang Zhao (2018): Less than 360° in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit der Studentin die Idee entwickelt, die sie implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir von der Studentin durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Da das Anschauen von Filmen für die meisten Menschen eine entspannende Tätigkeit ist, sind Techniken erforderlich, um den gesamten Bereich eines CVR-Films zu erkunden, ohne den Körper ständig drehen zu müssen. Solche Techniken lassen sich nicht nur auf dem Sofa, sondern auch am Schreibtisch, in einem Zug oder an anderen Orten mit eingeschränkter Bewegungsfreiheit anwenden.

Dafür wurden drei Methoden implementiert und miteinander verglichen:

const (Controller mit konstanter Rotationsgeschwindigkeit)

Mit der *const*-Methode kann der Benutzer das Bild mit konstanter Geschwindigkeit drehen, indem er die Controller-Taste nach rechts oder links drückt. Die Drehung ist nur in horizontaler Richtung möglich. Kopfbewegungen können zusätzlich zum Inspizieren der Szenerie verwendet werden. Für die Geschwindigkeit wurden 45°/s gewählt.

var (Controller mit variabler Rotationsgeschwindigkeit)

Die *var*-Methode ähnelt der *const*-Methode bezüglich der Drehtechnik, unterscheidet sich jedoch in der Geschwindigkeit. Je weiter die Teilnehmer die Taste nach rechts oder links betätigen, desto höher ist die Rotationsgeschwindigkeit bis zu einer maximalen Geschwindigkeit von 60°/s.

gain (Rotational Gain)

Für die *gain*-Methode wird die verstärkte Kopfbewegung von Jaekl genutzt (Jaekl u. a., 2005) (siehe Abschnitt 5.2.2). Die Ergebnisse von Jaekl zeigen, dass ein gain-Faktor (Verstärkungsfaktor) zwischen 0.88 und 1.33 für die Drehung des Kopfes von den Teilnehmern nicht als Modulation wahrgenommen wird. Unter Berücksichtigung eines Bereichs von 120° für maximale Kopfbewegungen und eines HMD-FoVs von 100° ist ein gain-Faktor von 1.33 nicht ausreichend, um den gesamten 360°-Bereich abzudecken.

Zur Ermittlung der geeignetsten Rotationsverstärkung wurden sechs gain-Faktor-Werte (1.3 - 1.5 - 1.7 - 1.8 - 2 - 2.1) in einer informellen Pilotstudie mit 5 Teilnehmern getestet. Daraus resultierend wurde der Faktor 2.0 gewählt. Dieser Gain-Faktor wird auch in den Untersuchungen von Jay und Hubbold genutzt (Jay und Hubbold, 2003): Ein *gain*-Faktor von 2 ist für den Betrachter angenehm, auch wenn die Verstärkung manchmal spürbar ist. Der Verstärkungsfaktor von 2.0 hat einen weiteren Vorteil. Selbst wenn sich Betrachter vollständig drehen, haben sie immer die gleichen Filmabschnitte an der gleichen Position, denn der Filminhalt wird in einer 360°-Runde genau zweimal dargestellt.

7.4.2.1 Material

Da wir an Situationen interessiert waren, in denen man zu Hause Filme schaut, wurden die Experimente als Feldstudie in einer häuslichen Umgebung durchgeführt. In der Studie sahen sich die Zuschauer die Filme über ein Google Cardboard mit einem Samsung Galaxy S7 auf einer Couch oder einem Stuhl (kein Drehstuhl) an. Das Cardboard wurde mit einem Kopfband versehen. Es war wichtig, dass kein Kabel die Bewegungen der Teilnehmer einschränkte.

Für die Implementierung der drei Methoden und die Aufzeichnung der Kopfbewegungen wurde Unity 2017.2 verwendet. Während der Wiedergabe von Filmen mit der Methode *const* und *var* benutzten die Teilnehmer einen auvisio mini Bluetooth-Controller.

In den Studien wurden drei Animationsfilme genutzt: „Unexpected Guest“ (Om Nom 360°, 2017), „Special Delivery“ (Aardman Animations, 2015) und ein Fragment von „The Dream Collector“ (Pinta Studios, 2017). Jeder Film dauerte ca. 2:30 min und nutzte den gesamten 360°-Bereich für die Filmhandlung.

7.4.2.2 Teilnehmer und Prozedur

18 Personen (5 weiblich, 13 männlich, Alter: MW=25.8 Jahre) nahmen an dem Experiment teil. 33.3 % von ihnen hatten keine VR-Erfahrung.

Es wurde ein within-subject Testdesign verwendet: Jeder Teilnehmer wendete alle drei Methoden an, jede mit einem anderen Film. Die Reihenfolge der Methoden und die Zuordnungen zwischen Film und Methode wurden durch ein Graeco-Latin-Quadrat ausbalanciert.

Nach einer kurzen Einführung sahen sich die Teilnehmer ein Übungsvideo an und testeten die drei Methoden. Der Fragebogen bestand aus mehreren Teilen: Demographie, Presence, Experience, Simulator-Sickness, Recall und am Ende ein direkter Vergleich der beiden Methoden.

7.4.2.3 Ergebnisse

Der Shapiro-Wilk-Test zeigte, dass nicht für alle Fragestellungen Normalverteilung angenommen werden konnte. Daher wurde der Friedman-Test angewendet, um signifikante Unterschiede zwischen den Methoden zu finden. Zur Bestimmung der Signifikanz wurde ein α -Level von 0.05 gewählt. Ein post-hoc Vorzeichen-Ranking Tests wurden zum paarweisen Vergleich der Methoden verwendet.

Presence

Der Friedman-Test ergab keine signifikanten Unterschiede zwischen den Methoden. Die Mittelwerte bewegten sich zwischen 4.06 und 5.22.

Experience

Drei der Experience-Fragen zeigten signifikante Unterschiede: EX1, EX3, EX6. In Bezug auf EX1 war die Punktzahl für die *gain*-Methode ($MW_{gain} = 5.33$, $SD_{gain} = 0.97$) höher als für die *var*-Methode ($MW^{var} = 4.67$, $SD_{var} = 1.08$) ($p_{gain/var} = 0.048$). Die *const*-Methode ($MW_{const} = 4.0$, $SD_{const} = 1.75$) wurde als nicht so einfach durchführbar bewertet (EX3) wie die beiden anderen Methoden ($p_{const/var} = 0.047$, $p_{const/gain} = 0.026$). Bei der *gain*-Methode ($MW_{gain} = 5.22$, $SD_{gain} = 1.11$) war es einfacher, mit dem Tempo des Inhalts (EX6) Schritt zu halten, als bei der *const*-Methode ($MW_{const} = 4.0$, $SD_{const} = 1.75$) ($p_{gain/const} = 0.02$). Die Mittelwerte sind in Abbildung 7-88 zu sehen.

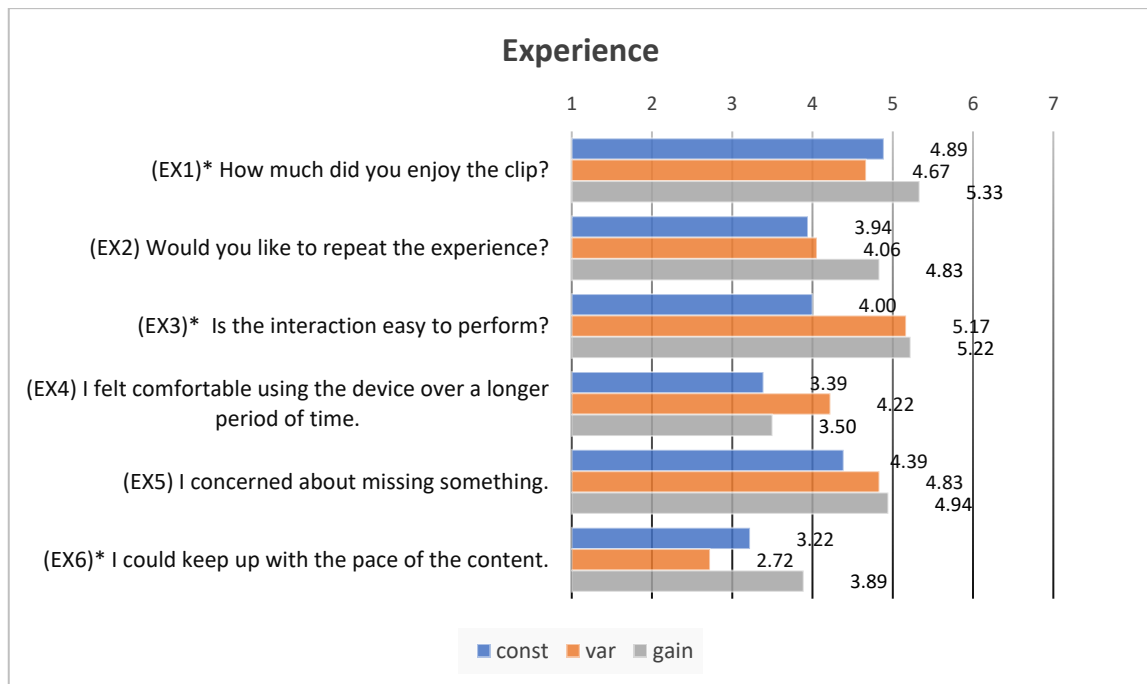


Abbildung 7-88: Mittelwert für die Experience-Items, die sich für EX1, EX3 und EX6 signifikant unterschieden (Likert-Skala von 1 bis 7).

Simulator-Sickness

Die Werte für den SSQ waren generell niedrig und lagen zwischen 0.11 und 1.11 (Tabelle 7-20). Für den Dizziness-Item (Schwindel) konnten signifikante Unterschiede festgestellt werden, der Friedman-Test ergab einen p-Wert von 0.009. Die *gain*-Methode zeigte die höchsten Werte: $MW_{gain} = 1.11$, $MW_{const} = 0.61$, $M_{var} = 0.39$. Paarweise post-hoc Tests ergaben signifikante Unterschiede zwischen der *gain*-Methode und den beiden *controller*-Methoden: $p_{gain/var} = 0.003$, $p_{gain/const} = 0.046$.

	general discomfort	headache	dizziness	eye strain	difficulty focusing
const	0.78	0.28	0.61	0.94	1.11
var	0.89	0.11	0.39	0.67	0.61
gain	0.72	0.33	1.11	1.11	0.94

Tabelle 7-20: Bei den SSQ-Items gibt es signifikante Unterschiede bei Schwindelgefühl (dizziness). Signifikant unterschiedliche Werte sind fett dargestellt.

Recall

Nach jedem Video wurden drei Fragen zum Videoinhalt gestellt. Bei all diesen Fragen wurden visuelle Aspekte aber kein Audio-Inhalt berücksichtigt. Für jede richtige Antwort wurde ein Punkt vergeben. Die maximale Punktzahl pro Video und Teilnehmer war drei.

Die Recall-Fragen resultierten in den folgenden Scores: $MW_{const} = 1.72$, $SD_{const} = 0.86$; $M_{var} = 1.78$, $SD_{var} = 0.77$; $MW_{gain} = 1.42$, $SD_{gain} = 0.8$. Es gab keine signifikanten Unterschiede zwischen den Methoden ($p = 0.37$).

Vergleich

Nachdem alle drei Methoden angewendet wurden, verglichen die Teilnehmer diese direkt miteinander. Für die Antworten sollte eine kurze Erklärung gegeben werden. Nur 11.1 % der Teilnehmer bewerteten die *const*-Methode als die bequemste, 38.9 % die *gain*-Methode und 50 % die *var*-Methode (Abbildung 7-89).

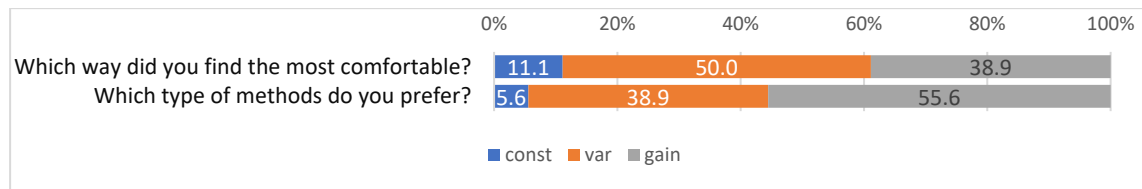


Abbildung 7-89: Die meisten Zuschauer bevorzugten die *var*- und die *gain*-Methode und fanden diese auch komfortabler als die *const*-Methode.

Die von den Teilnehmern angegebenen Gründe betrafen häufig die Bequemlichkeit, sowohl für die *var*-Methode (P3, P5, P10, P11, P14, P18) als auch für die *const*-Methode (P6) und die *gain*-Methode (P12).

Im Gegensatz zu den Ergebnissen der Presence-Fragebögen gaben mehrere Teilnehmer an, dass sie sich mit der *gain*-Methode am wohlsten fühlen, weil sie realistischer ist (P1, P4, P7, P8, P9). Weitere Vorteile, die erwähnt wurden:

Vorteile des Controllers in der *const* – und *var*-Methode:

- „bequem“ (P13)
- „Ich muss meinen Kopf nicht drehen“ (P4, P18)
- „Man kann sich auf das Zuschauen konzentrieren und muss sich nicht umsehen“ (P17)

Vorteile der *gain*-Methode:

- „Ich konnte den Winkel und die Geschwindigkeit kontrollieren“ (P16)
- „realer“ (P1, P4, P7, P8, P9)
- „kein Schwindel“ (P7)
- „konnte mit dem Kopf gesteuert werden“ (P16)

7.4.2.4 Diskussion

Die Ergebnisse der Studie zeigen, dass die *var*- und die *gain*-Methode von den Zuschauern akzeptiert wurden. Die *const*-Methode ergab die niedrigsten Scores, insbesondere im Vergleichsteil. Die Methoden haben Vor- und Nachteile. Die *gain*-Methode erhielt höhere Scores für die Experience, verursachte jedoch mehr Simulator-Sickness. Die Ergebnisse der Fragebogenteile sind in Tabelle 7-21 zusammengefasst.

	positiv	negativ
Presence	-	-
Experience	gain	const
Simulator-Sickness	var	gain
Vergleich	var /gain	const

Tabelle 7-21: Zusammenfassung der Versuchsergebnisse für jeden Fragebogenteil.

Mit der *gain*-Methode kann der gesamte 360°-Bereich angezeigt werden, indem der Kopf in einem 180°-Bereich gedreht wird. Es ist jedoch schwierig zu erkennen, dass die Dinge auf der linken Seite zu denen auf rechten Seite gehören. Da es keine Begrenzung auf der linken und rechten Seite gibt, scheint es, als würde die Szene dort fortgesetzt, obwohl es in nur eine Wiederholung der Szene ist. Elemente, die für den Betrachter weit auseinander liegen, gehören eigentlich zusammen. Diese Tatsache wurde von den Zuschauern nicht beanstandet oder erwähnt. Wir haben jedoch nicht untersucht, ob die Teilnehmer davon überhaupt Kenntnis genommen haben und ob die räumliche Orientierung beeinträchtigt war.

Die Methoden sind für Filme anwendbar, bei denen es wichtig ist, den gesamten 360°-Bereich gut zu sehen, die räumliche Orientierung aber weniger relevant ist. Es stellt sich die Frage, ob der gesamte 360°-Bereich für ein umfassendes CVR-Filmerlebnis tatsächlich erforderlich ist und aufgezeichnet werden muss. Wir werden das in **Studie 7.4.4** untersuchen.

In unserer Studie zur Rotationsverstärkung wurde die Kopfdrehung nur in horizontaler Richtung verstärkt. Es ist jedoch auch möglich, dies für die vertikalen Bewegungen umzusetzen, um ein kleines FoV auszugleichen und es komfortabler zu machen, nach oben und unten zu schauen.

Aus den oben genannten Gründen (am bequemsten, weniger Simulator-Sickness) wurde die *var*-Methode aus dieser Studie ausgewählt und im nächsten Experiment mit zwei anderen Methoden verglichen.

7.4.3 Studie: Vergleich der Rotationsmedien

Die Studie ist Teil der Masterarbeit von Lang Zhao (2018): Less than 360° in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit der Studentin die Idee entwickelt, die sie implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir von der Studentin durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

In dieser Studie wurden die folgenden drei Methoden betrachtet:

fixed (Rotation des Kopfes – feste Sitzposition)

Der Zuschauer sitzt auf einem normalen Stuhl und kann sich umsehen, indem der Kopf bewegt wird. Die Teilnehmer wurden darüber informiert, dass die Handlung des Films vor ihnen abläuft und sich keine wichtigen Details der Geschichte hinter ihnen abspielen.

swivel (Rotation des Körpers mittels eines Drehstuhls)

Der Zuschauer sitzt auf einem Drehstuhl und kann sich in alle Richtungen drehen. Außerdem kann der Kopf bewegt werden, um sich umzusehen.

controller (Rotation des Films)

Die *controller*-Methode mit variabler Geschwindigkeit aus **Studie 7.4.2** (*var*) wurde wiederverwendet: Der Benutzer kann das Bild mit dynamisch angepasster Geschwindigkeit drehen, indem er die Controller-Taste nach rechts oder links drückt. Je mehr die Taste betätigt wird, desto höher ist die Rotationsgeschwindigkeit bis zu einer maximalen Geschwindigkeit von 60°/s

7.4.3.1 Material

Aus den gleichen Gründen wie in **Studie 7.4.2** verwendeten wir ein Google-Cardboard und ein Samsung Galaxy S7. Für die Filmrotation wurde ein auvisio mini Bluetooth-Controller und das Skript der **Studie 7.4.2** genutzt.

Es wurde ein Film ausgewählt, in dem die meisten Aktionen vor dem Benutzer in einem Bereich von 225° stattfinden und die Filmhandlung allein durch Drehen des Kopfes erfasst werden kann: „Son of Jaguar“ (Gutierrez, 2017). Der Teil hinter dem Zuschauer wurde nicht gelöscht. Die Teilnehmer wurden jedoch darauf hingewiesen, dass sich dort keine story-relevanten Details abspielen. Der 9-minütige Film wurde in drei 3-minütige Teile unterteilt.

7.4.3.2 Teilnehmer und Prozedur

24 Personen (10 weiblich, 14 männlich, Alter: MW=25.25) nahmen an dem Experiment teil. 37.5 % von ihnen hatten keine VR-Erfahrungen. Dabei wurde ein within-subject Testdesign verwendet. Jeder Benutzer sah sich alle drei Teile des Films mit einer anderen Methode an: einen Teil auf einem Drehstuhl, einen auf einem normalen Stuhl und einen auf einem normalen Stuhl mit einem Controller. Die Reihenfolge der Filmteile war gleich, die Reihenfolge der Methoden wurde permutiert. Alle sechs Permutationen wurden für die gleiche Anzahl von Teilnehmern verwendet.

Der Fragebogen zu **Studie 7.4.2** wurde überarbeitet und an diese Studie angepasst. Die Aufmerksamkeitsfragen wurden auf den neuen Filminhalt transformiert und die Vergleichsfragen geändert in:

- (C1) Welche Methode fanden Sie am bequemsten?
- (C2) Welche Art von Videos würden Sie bevorzugen? (180°/360°)
- (C3) Hatten Sie das Bedürfnis, ihren Kopf oft zu drehen?

Auch hier sollten die Teilnehmer die Antworten begründen. Zusätzlich wurden die Kopfbewegungen jedes Zuschauers aufgezeichnet.

7.4.3.3 Ergebnisse

Es wurden der Friedman-Test und paarweise post-hoc Vorzeichen-Ranking-Tests verwendet, um signifikante Unterschiede zwischen den Methoden zu finden.

Presence

Bezüglich der Presence-Fragen konnten wir keine signifikanten Unterschiede feststellen.

Experience

Auch bei den Experience-Fragen waren keine signifikanten Unterschiede zwischen den Methoden erkennbar.

Simulator-Sickness

Es gab nur wenig Simulator-Sickness (Mittelwerte zwischen 0.25 und 1.00). Der Friedman-Test ergab keine signifikanten Unterschiede zwischen den Methoden (Tabelle 7-22).

	general discomfort	headache	dizziness	eye strain	difficulty focusing
fixed	1.00	0.38	0.75	0.75	0.88
swivel	0.88	0.25	1.00	0.54	0.67
controller	0.79	0.33	0.88	0.58	0.83

Tabelle 7-22: Mittelwerte der SSQ-Items.

Vergleich

45.83 % der Teilnehmer fühlten sich mit der *controller*-Methode am wohlsten, 41.67 % mit der *swivel*-Methode und 12.5 % mit der *fixed*-Methode (Abbildung 7-90). Trotzdem würden 79.17 % der Teilnehmer 180°-Filme den 360°-Filmen vorziehen (Abbildung 7-91).

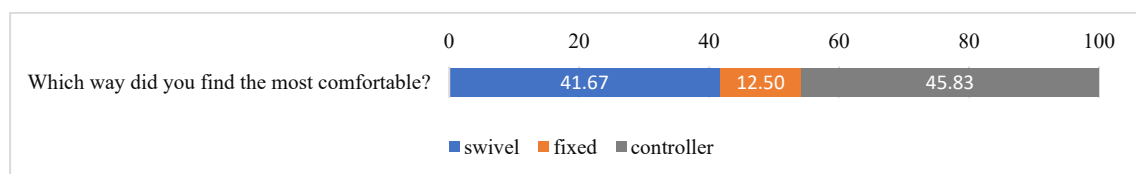


Abbildung 7-90: Die wenigsten Zuschauer fanden die fixed-Methode am komfortablesten (Antworten in %).

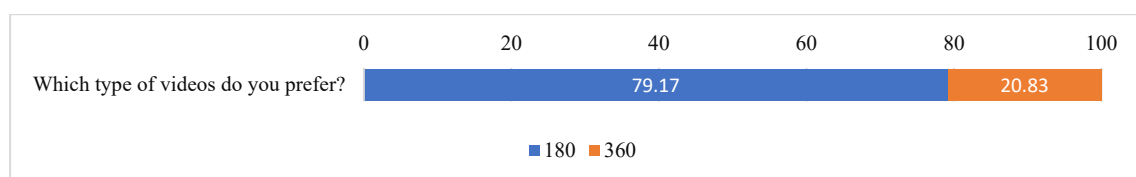


Abbildung 7-91: Die meisten Teilnehmer würden lieber 180° Filme schauen (Antworten in %).

Die Frage, ob die Teilnehmer ihren Kopf oft drehen möchten (C3) wurde auf einer 7-Punkte-Likert-Skala beantwortet (1 = Ja, ich erkundete und sah gern neue Räume, 7 = Nein, ich möchte nicht häufig den Kopf drehen, ich möchte es nur bequem genießen) mit einem Mittelwert von 4.9 (SD = 1.8) beantwortet.

Analyse der Tracking-Daten

Die Headtracking-Daten wurden für alle drei Methoden und Filme separat geprüft, um Unterschiede zwischen den Testfällen festzustellen. Mit dem CVR-Analyzer wurden die spatio-temporalen Daten sowohl für den gesamten Film wie auch für Teile des Films untersucht.

Die Muster für die *swivel*- und die *fixed*-Methode waren größtenteils sehr ähnlich (Abbildung 7-92). Die mit der *controller*-Methode erzeugten Visualisierungen zeigten weniger Konzentration und kleinere verstreute Heatmap-Bereiche. Ein Grund dafür könnte sein, dass es schwieriger ist, Details zu fixieren und sich mühelos im Bild zu bewegen. Außerdem kann es komplizierter sein, einen zuvor inspizierten Bereich wiederzuentdecken. Für die *fixed*-Methode und auch für den Drehstuhl kann das menschliche Orientierungssystem die Rückkehr zu einem zuvor gesehenen Bereich unterstützen, um dort weiter der Handlung zu folgen.

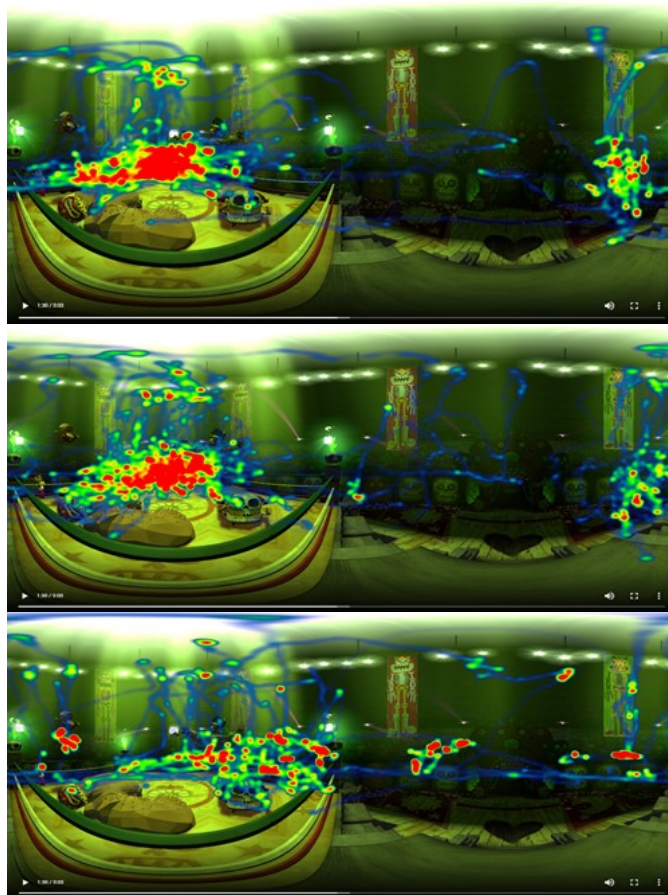


Abbildung 7-92: Heatmaps für den 2. Film: *swivel* (oben), *fixed* (Mitte), *controller* (unten).

7.4.3.4 Diskussion

Die Methoden unterschieden sich in Bezug auf Presence, Erlebnis und Simulator-Sickness nicht signifikant. Diese Ergebnisse legen nahe, dass die Verwendung eines Drehstuhls zum Ansehen von CVR-Filmen nicht unbedingt notwendig ist und durch andere Methoden ersetzt werden könnte.

Selbst wenn 87.5 % der Teilnehmer die Rotationstechniken (*controller* oder *swivel*) am angenehmsten empfanden, würden die meisten Teilnehmer 180°-Filme bevorzugen. Die Tatsache, dass wir die Rückseite des Films nicht gelöscht haben, könnte das Ergebnis beeinflusst haben, dass die *fixed*-Methode schlechter abschnitt.

In den spatio-temporalen Daten fanden wir ähnliche Muster für die *swivel*- und die *fixed*-Methode: relativ stabile Bereiche um die Rols. Bei der *controller*-Methode waren die Muster diffuser, es scheint schwieriger zu sein eine Region zu fokussieren und zu einem zuvor inspizierten Bereich zurückzukehren.

7.4.4 Studie: Weniger als 360°

Die Studie ist Teil der Bachelorarbeit von Arne Fahrenwalde (2019): *Less than 360° in Cinematic Virtual Reality*.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Da viele Zuschauer nicht ständig den Kopf drehen möchten und 180°-Videos den 360°-Videos vorziehen würden (**Studie 7.4.3**), wurden in dieser Studie 180°- und 225°-Videos mit der *swivel*-Methode, verglichen, die üblicherweise für den gesamten 360°-Videos verwendet wird. Das Ziel dieses Experiments war es, zu untersuchen, ob der reduzierte Bereich die Presence, Erlebnis oder Simulator-Sickness beeinflusst.

Es wurden drei Methoden miteinander verglichen:

180° (fester Stuhl)

Der reduzierte Bereich von 180° wurde gewählt, da dieses Format bereits von Google und YouTube eingeführt wurde. Die Grenzen dieses Bereichs wurden mit einem 4.5°-breiten Rand ausgeblendet. Wenn die Teilnehmer geradeaus blickten, sahen sie diese Grenzen nicht. Horizontale Kopfbewegungen sind für Menschen in einem Bereich von bis zu 45° in jede Richtung angenehm (Dreyfuss, 1967). Unter Berücksichtigung des 100° FoV der Samsung Gear (SM-R323) konnten die Teilnehmer somit leicht einen Bereich von 190° und damit die Grenzen erkennen, wenn sie ihren Kopf nach links oder rechts bewegten. Die Teilnehmer sahen sich den Film in einer festen Sitzposition an und bewegten nur den Kopf.

225° (fester Stuhl)

Der 225°-Bereich wurde aufgrund des FoV des Samsung Gear HMDs (100°) und der maximalen Kopfdrehung des Menschen (60° in jede Richtung) gewählt. Die Grenzen wurden über einen 4.5°-breiten Rand ausgeblendet. Bei bequemen Kopfbewegungen (45° in jede Richtung) waren die Grenzen nicht sichtbar. Die Grenzen waren jedoch bei extremen Kopfbewegungen zu sehen. Wie im 180°-Fall sahen die Teilnehmer den Film in einer festen Sitzposition und bewegten nur den Kopf.

360° (Drehstuhl)

Bei der dritten Methode konnten die Betrachter den gesamten 360°-Bereich mit einem Drehstuhl inspizieren.

7.4.4.1 Material

Die Teilnehmer sahen sich die Filme mittels eines Samsung Gear-Headset und eines Samsung S7-Smartphones an. Es wurden drei Filme ausgewählt, die sich in Typ, Länge und Tempo unterscheiden: professionelle Dokumentarfilme, bei denen die Hauptaktion in einem 180°-Bereich stattfindet: „Lions 360°“ (National Geographic, 2017a), „Galapagos 360°“ (National Geographic, 2018), „360°Underwater National Park“ (National Geographic, 2017b). Zwei der drei Videos wurden bearbeitet, damit alle Filme eine ähnliche Länge haben (ca. 3 Minuten) und die Aktionen immer vor dem Betrachter ablaufen.

7.4.4.2 Teilnehmer und Prozedur

24 Personen (12 weiblich, 12 männlich, Alter: MW=35.79) nahmen an dem Experiment teil. Eine Person hatte keine VR-Erfahrungen, 17 nur wenig. Es wurde ein within-subject Testdesign verwendet. Jeder

Benutzer schaute sich drei verschiedene Filme an, jeden mit einer der Methoden. Filme und Methoden wurden mit einem Graeco-Latin-Quadrat ausbalanciert.

Der Fragebogen ähnelte dem der **Studie 7.4.3**. Es wurden vier Fragen des NASA-TLX hinzugefügt, um die Belastung für die verschiedenen Methoden zu vergleichen.

7.4.4.3 Ergebnisse

Da der Shapiro-Wilk-Test zeigte, dass nicht für alle Items Normalverteilung angenommen werden konnten, wurden der Friedman-Test und der paarweise signed-ranks Test angewendet, um signifikante Unterschiede zwischen den Methoden zu finden (α -Level 0.05).

Presence

Der Friedman-Test ergab mehrere signifikante Unterschiede. Für den Punkt „Ich hatte das Gefühl da zu sein“ (1 = stimme überhaupt nicht zu, 7 = stimme voll und ganz zu) war der Score für die 360°-Methode ($MW_{360} = 5.88$, $SD_{360} = 1.19$) signifikant höher als für die 180°-Methode ($MW_{180} = 4.75$, $SD_{180} = 1.39$, $p_{360/180} < 0.01$) und die 225°-Methode ($MW_{225} = 5.29$, $SD_{225} = 1.2$, $p_{360/225} = 0.07$). Zusätzlich hatten im 360°-Film mehr Teilnehmer das Gefühl, dass die virtuelle Welt sie umgibt ($MW_{360} = 6.21$, $SD_{360} = 1.25$) als im 180°-Film ($MW_{180} = 3.33$, $SD_{180} = 1.88$, $p_{360/180} < 0.01$) und im 225°-Film ($MW_{225} = 3.79$, $SD_{225} = 1.93$, $p_{360/225} < 0.001$). Die Scores sind in Abbildung 7-93 dargestellt.

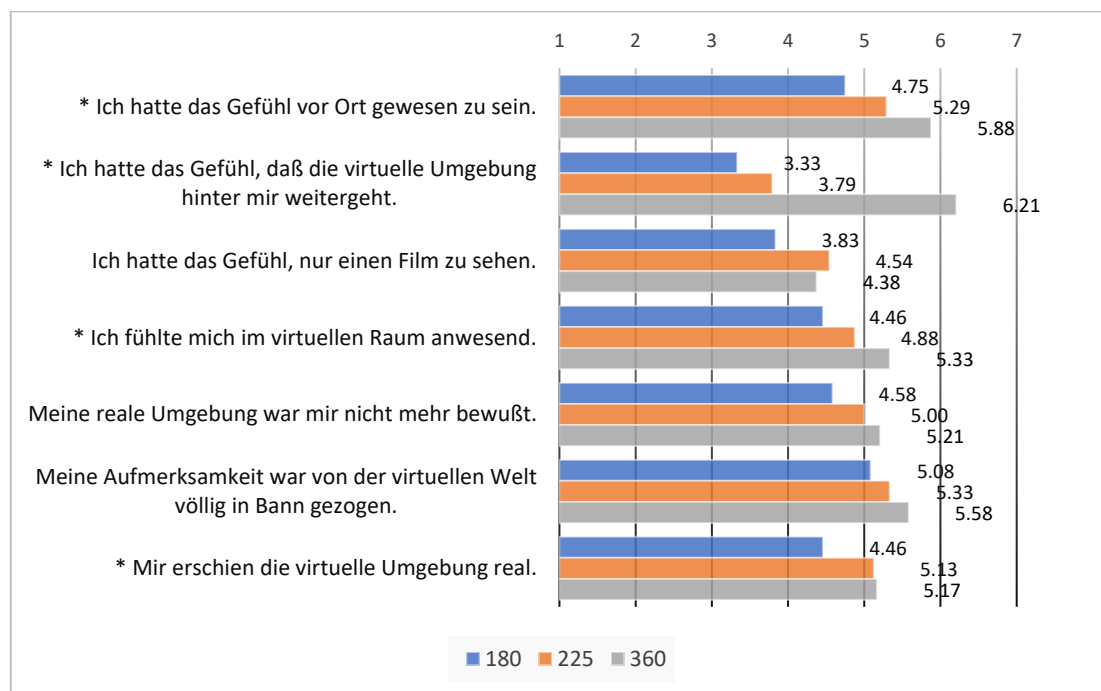


Abbildung 7-93: Mittelwerte für die Presence-Items (Likert-Skala von 1 bis 7). Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

Task Workload

Die NASA-TLX-Ergebnisse ergaben höhere Scores für die 360°-Bedingung bezüglich mehrerer Items.

Die physische Belastung war bei der 360°-Methode ($MW_{360} = 4.71$, $SD_{360} = 2.51$) signifikant höher als bei der 180°-Methode ($MW_{180} = 2.46$, $SD_{180} = 1.67$) ($p_{360/180} < 0.01$) und der 225°-Methode ($MW_{225} = 3.08$, $SD_{225} = 1.84$, $p_{360/225} = 0.02$).

Auch die zeitliche Belastung war bei der 360°-Methode ($MW_{360} = 4.58$, $SD_{360} = 2.75$) höher als bei der 180°-Methode ($MW_{180} = 2.63$, $SD_{180} = 2.28$, $p_{360/180} < 0.01$) und der 225°-Methode ($MW_{225} = 2.96$, $SD_{225} = 2.2$, $p_{360/225} = 0.03$).

Die Mittelwerte sind in Abbildung 7-94 dargestellt.

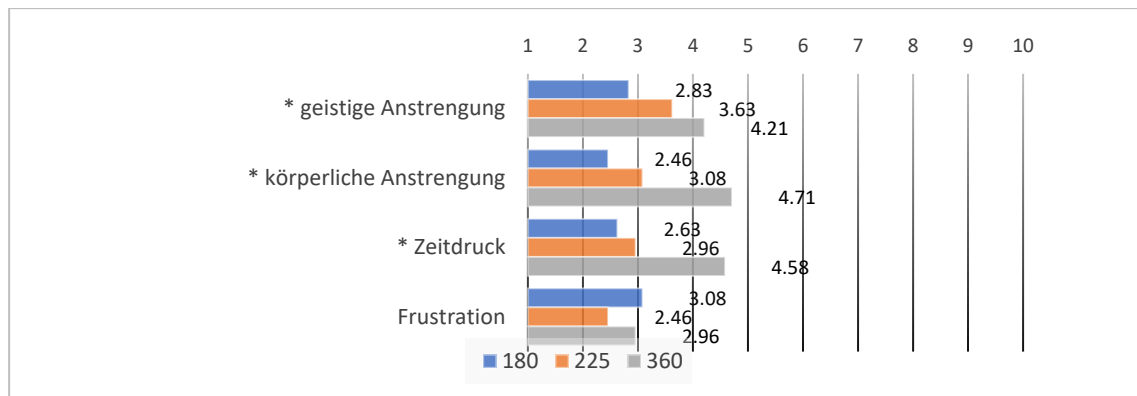


Abbildung 7-94: Mittelwerte für die Task Workload Items (Likert-Skala von 1 bis 10). Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

Erlebnis

Der Punkt „Ich habe das Filmerlebnis genossen.“ (1 = überhaupt nicht, 7 = sehr) ergab die höchste Punktzahl für die 360°-Methode ($MW_{360} = 5.92$, $SD_{360} = 1.14$, $p = 0.02$) und die niedrigste für die 180°-Methode ($MW_{180} = 5.17$, $SD_{180} = 1.17$). Im 225°-Fall hatten weniger Teilnehmer „Angst, etwas zu verpassen“ ($MW_{225} = 2.3$, $SD_{225} = 1.4$) als im 360°-Fall ($MW_{360} = 4.2$, $SD_{360} = 1.8$) ($p_{225/360} = 0.016$). Die Mittelwerte sind in Abbildung 7-95 dargestellt.

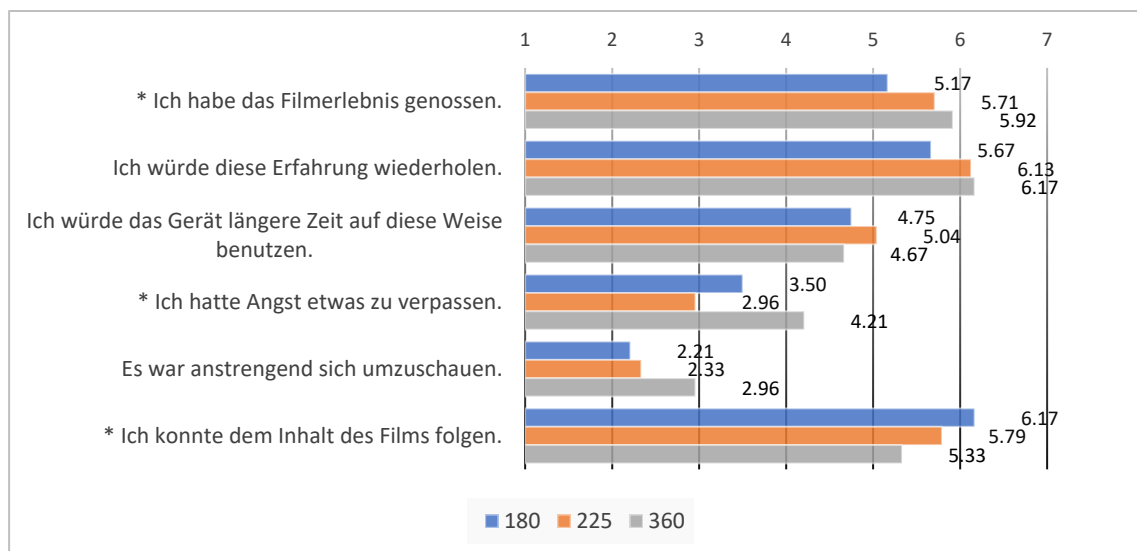


Abbildung 7-95: Mittelwerte für die Erlebnis-Items (Likert-Skala von 1 bis 7). Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

Simulator-Sickness

Die Werte für die SSQ waren im Allgemeinen sehr niedrig und lagen zwischen 0.17 und 0.5 (Tabelle 7-23). Es gab keine signifikanten Unterschiede zwischen den Methoden.

	Allgemeines Unwohlsein	Ermüdung	Kopfschmerz	Überanstrengte Augen
180°	0.38 (0.58)	0.29 (0.55)	0.17 (0.38)	0.5 (0.72)
225°	0.21 (0.59)	0.21 (0.51)	0.17 (0.38)	0.5 (0.59)
360°	0.38 (0.71)	0.38 (0.65)	0.21 (0.41)	0.5 (0.72)

Tabelle 7-23: Mittelwerte und Standardabweichungen (in Klammern) für die SSQ Items.

Vergleich

Nachdem die Teilnehmer alle drei Videos gesehen hatten, beurteilten sie auf einer 7-Punkte-Likert-Skala, wie sehr ihnen jede Methode gefallen hatte (1 = überhaupt nicht, 7 = sehr). Die Werte für die Methoden 225° ($MW_{225}=5.6$, $SD_{225}=1.2$) und 360° ($MW_{360}=5.5$, $SD_{360}=1.4$) sind ähnlich und sind signifikant höher als der Wert der Methode 180° ($MW_{180}= 4.8$, $SD_{180}= 1.5$, $p_{180/225}=0.03$, $p_{180/360}=0.5$).

Die Teilnehmer hatten die Möglichkeit, zu jeder Methode Anmerkungen zu notieren, so dass wir Einblicke in Vor- und Nachteile bekommen konnten.

Vorteile von 360°:

- „Gefühl, mitten in der Szene zu sein“ (P6)
- „sich frei umsehen“ (P6)
- „realistischer“ (P4)

Nachteile von 360°:

- „ein starkes Gefühl der Überforderung, da viel passiert ist“ (P13)
- „Man weiß nicht, wo die Haupthandlung ist und wohin man sich drehen soll.“ (P15)
- „anspruchsvoll“ (P4)
- „Drehen ist unangenehm.“ (P8)

Vorteile von 225°:

- „Man ist nicht verloren, da man weiß, wo die Hauptaktion ist.“ (P15)

Nachteile von 225°:

- „Man wusste, dass man sich nicht vollständig umsehen kann.“ (P15)

Vorteile von 180°:

Die Teilnehmer nannten keine Vorteile. Die Bemerkung für 225°, dass die Zuschauer weniger verloren sind, gilt jedoch auch für 180°.

Nachteile von 180°:

- „Gefühl, eingeschränkt zu sein“ (P6)
- „zu kleines Sichtfeld“ (P15)
- „eher wie traditionelles Fernsehen“ (P1)

Analyse der Tracking-Daten

Die Tracking-Daten wurden für alle drei Methoden analysiert. Auffällig ist, dass die Tracks und Heatmaps für den 180°- und den 225°-Fall sehr ähnlich sind (Abbildung 7-96).

Die Teilnehmer, die den 225°-Film sahen, drehten nicht mehr den Kopf als die Zuschauer des 180°-Films. Der inspizierte Bereich (mit Ausnahme einiger Ausreißer) hat für beide eine horizontale Ausdehnung von etwa 145°. Da wir uns für ein komfortables Consumer-Gerät ohne Kabel entschieden hatten, sind keine Eyetracking-Daten verfügbar. Es ist jedoch zu erwarten, dass im 180°-Fall die Grenzen präsenter waren als im 225°-Fall. Überraschenderweise drehten die Betrachter den Kopf unter beiden Bedingungen um mehr als die 120°, die als reine Kopfdrehung angenommen wurde. Um mehr zu sehen, wurde auch der Oberkörper leicht gedreht. Es fällt auch auf, dass die Betrachter im 180°- und im 225°-Fall mehr die vertikale Richtung inspizierten, während der Heatmap Bereich für den 360°-Film flacher ist.

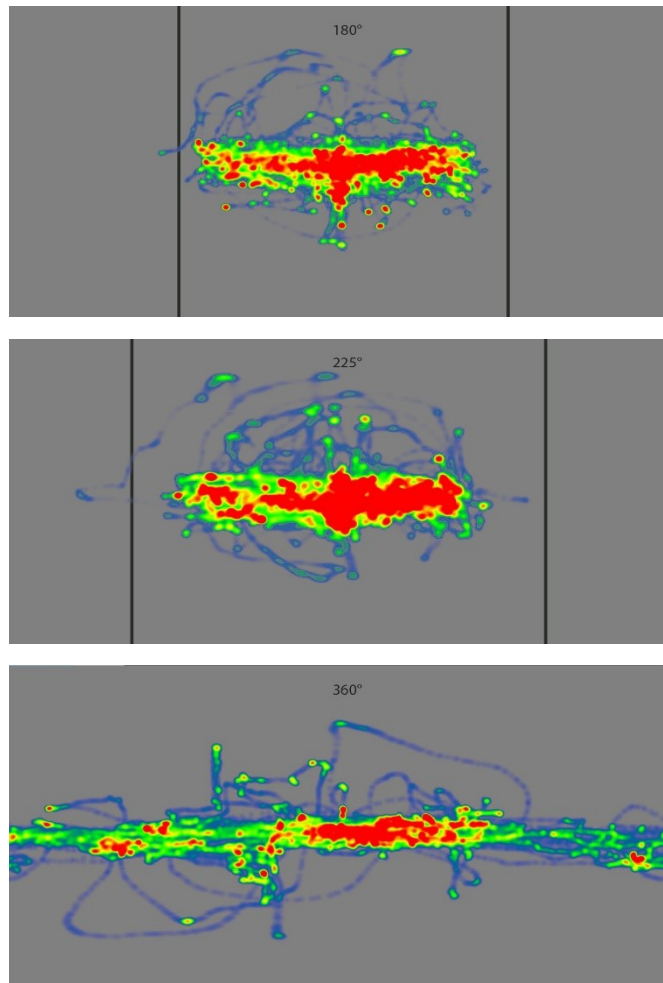


Abbildung 7-96: Heatmaps für einen der Filme mit allen drei Methoden (180° oben, 225° Mitte, 360° unten). Die Heatmaps für 180° und 225° sind sehr ähnlich.

7.4.4.4 Diskussion

Selbst wenn die Zuschauer auf einem festen Stuhl saßen, bei dem der Betrachtungsbereich normalerweise nicht größer als 225° ist (Kopf und Blick), verringerte sich die Presence geringfügig, wenn nicht der gesamte 360°-Bereich verfügbar war. Allein das Wissen, dass der Inhalt nicht die gesamte Umgebung abdeckt, mindert das Presence-Gefühl. Andererseits nahm aber die Belastung ab. Die Ergebnisse der Fragebogenteile sind in Tabelle 7-24 zusammengefasst.

	positiv	negativ
Presence	360°	180°
Erlebnis	360°/225°/180°	-
Simulator-Sickness	-	-
Task Workload	180°	360°
Vergleich	360°/225°	180°

Tabelle 7-24: Zusammenfassung der Versuchsergebnisse. für jeden Fragebogenteil.

In unserem Experiment wurde die Samsung Gear mit einem FoV von 100° verwendet. Für andere HMDs ist dieser Bereich größer und wird mit dem technologischen Fortschritt zunehmen. Ein 225°-Film könnte also in Zukunft zu klein für ein uneingeschränktes CVR-Filmerlebnis sein. Da das maximale FoV der menschlichen Augen etwa 188° (Dreyfuss, 1967) und die maximale Kopfbewegung 120° beträgt, kann ein

Film von etwa 308°-312° erforderlich sein. Auch in diesem Fall verbleibt ein kleiner Bereich für das Filmteam, um die Aufnahmebedingungen zu verbessern. Die Reduzierung des Inhaltsbereichs von Filmen ist jedoch auch wichtig, um den Betrachter nicht zu überlasten und um ihm die Befürchtung zu nehmen, etwas zu verpassen. Es bedarf zusätzlicher Untersuchungen, um die perfekten Parameter für alle HMDs zu finden.

7.4.5 Studie: Vergleich dynamischer und statischer Untertitel

Die Studie ist Teil der Bachelorarbeit von Kim Tran (2017): Subtitles in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit dem Studenten die Idee entwickelt, die er implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir vom Studenten durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Die Ergebnisse dieser Studie wurden vorab veröffentlicht in:

- Rothe, Sylvia, Tran, K., Hussmann, H., 2018. Positioning of Subtitles in Cinematic Virtual Reality, in: International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments - ICAT-EGVE 2018
- Rothe, Sylvia, Tran, K., Hußmann, H., 2018. Dynamic Subtitles in Cinematic Virtual Reality, in: Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '18. ACM Press, New York, USA

In dieser Studie wurden screen-referenzierte und world-referenzierte Untertitel miteinander verglichen.

screen-referenzierte Untertitel

Der Text ist statisch mit dem HMD verbunden. Unabhängig von der Blickrichtung sind die Untertitel mittig, 12.5° unter der Mittellinie platziert. Zur Sprecheridentifikation ist der Name des Sprechers am Anfang des Textes hinzugefügt.

world-referenzierte Untertitel

Der Text mit der Filmwelt verbunden und in der Nähe des Sprechers positioniert. Es hängt vom Szenario ab, wo die Untertitel genau platziert sind.

7.4.5.1 Material

Die Teilnehmer sahen zwei kurze Szenen, die in einem Fernsehstudio aufgenommen wurden (Gesamtlänge 3 Minuten). Dafür wurde eine Oculus Rift genutzt. In der ersten Szene (Abbildung 7-97) unterhalten sich zwei Personen, die ihren Platz nicht ändern. Wir nennen dieses Video *talk*-Video (T).

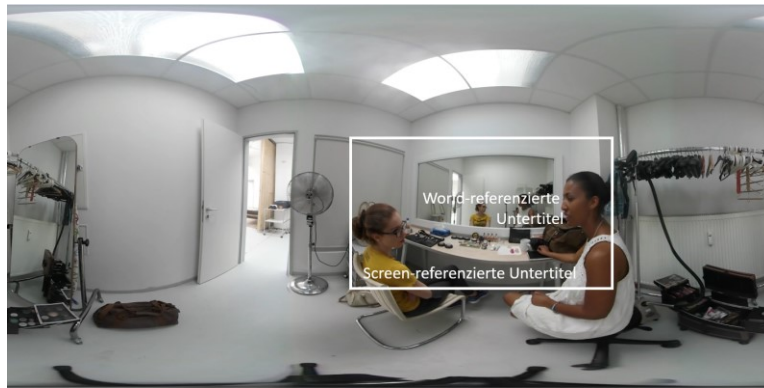


Abbildung 7-97: Die Szene des *talk*-Videos (T): Es gibt zwei sprechende Personen. Der Rahmen zeigt das Sichtfeld, der Text die Positionen der Untertitel (Schrift vergrößert), oben: world-referenziert, unten: screen-referenziert.

In der zweiten Szene (Abbildung 7-98) befinden sich mehrere Personen in einem Besprechungsraum, andere kommen und gehen. Dieses Video wird als *meeting*-Video bezeichnet (M). Um sicherzustellen, dass die Teilnehmer den gesprochenen Text nicht verstehen und auf die Untertitel angewiesen sind, wurde der Ton in beiden Videos manipuliert.



Abbildung 7-98: Die Szene des *meeting*-Videos (M): Mehrere Personen im selben Raum. Der Rahmen zeigt das Sichtfeld, der Text die Positionen der Untertitel (Schrift vergrößert), oben: world-referenziert, unten: screen-referenziert.

7.4.5.2 Teilnehmer und Prozedur

34 Teilnehmer (26 männlich, 8 weiblich, Alter: MW=22.9) sahen sich die Videos mittels HMD an. 23 Teilnehmer hatten etwas Erfahrung in VR, 19 verwendeten Untertitel in ihrem täglichen Leben. Falls Teilnehmer keine Erfahrungen mit VR hatten, wurde ihnen eine Einführung gegeben und vor der Studie ein kurzer Film ohne Untertitel gezeigt. Der Fragebogen bestand aus mehreren Teilen: Demographie, Task Workload, Simulator-Sickness, Presence, Vergleich der beiden Methoden.

Jeder Teilnehmer sah die gleichen zwei kurzen Videos, jedes mit einer der beiden Methoden. Die Reihenfolge der Videos und Methoden wurde unter Verwendung aller vier möglichen Kombinationen ausbalanciert:

- Tw - Gespräch (Talk) mit world-referenzierte Untertiteln
- Ts - Gespräch (Talk) mit screen-referenzierte Untertiteln
- Mw - Meeting mit world-referenzierten Untertiteln
- Ms - Meeting mit screen-referenzierten Untertiteln

So wurde jede Video- / Methodenkombination von 17 Teilnehmern angesehen. Alle Kopfbewegungen wurden aufgezeichnet.

Nach jedem Video wurden per Fragebogen Simulator-Sickness, Presence und Task Workload ermittelt und ein semistrukturiertes Interview geführt und aufgezeichnet. Dieses Interview basierte auf den Fragen von Brown et al. (Brown u. a., 2015) mit einigen ergänzenden Fragen zum Filminhalt.

7.4.5.3 Ergebnisse

In unseren Experimenten wurden zwei Arten von Daten erhoben: die Headtracking-Daten (spatio-temporale Daten) und die Antworten der Fragebögen.

Analyse der Tracking-Daten

Für die Analyse der Headtracking-Daten nutzten wir den von uns implementierten CVR-Analyzer und erstellten Heatmaps für jede der Szenen mit den aufgenommenen Daten jeder Methode. Für das Meeting-Video zeigten die Daten ein ähnliches Verhalten der Kopfbewegungen bei der Verwendung der beiden Methoden. Die Hotspots waren im Raum verteilt - nicht nur um die sprechenden Personen.

Im Dialog des *talk*-Videos waren die Teilnehmer jedoch stärker auf die Personen fokussiert, wenn die world-referenzierte Methode verwendet wurde (Abbildung 7-99).

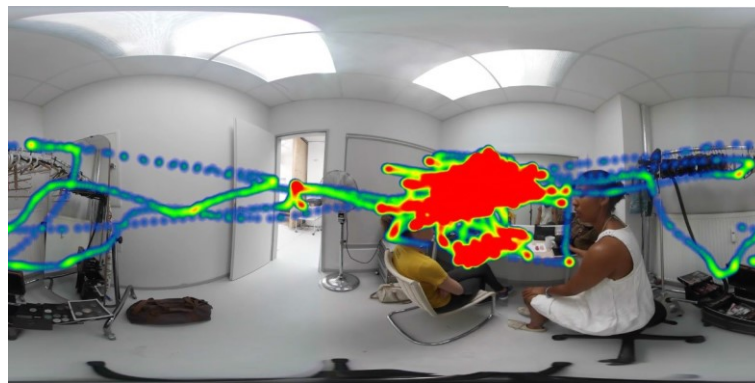


Abbildung 7-99: Heatmap der Headtracking-Daten für das *talk*-Video mit world-referenzierten Untertiteln (Tw). Im Bereich der Untertitel befindet sich ein Hotspot.

Die Teilnehmer, die dieselbe Szene mit der screen-referenzierten Methode betrachteten, sahen sich während des Dialoges der Protagonisten häufiger um (Abbildung 7-100).

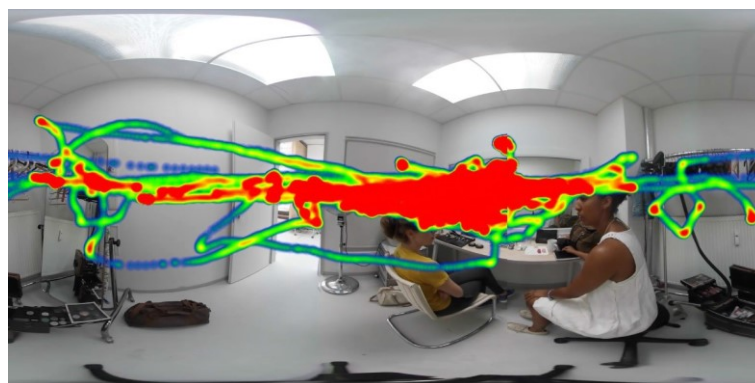


Abbildung 7-100: Heatmap der Headtracking-Daten für das *talk*-Video mit screen-referenzierten Untertiteln (Ts). Die Zuschauer schauten sich mehr um als in dem world-referenzierten Fall.

Task Workload

Die Scores für die meisten Items waren für die screen-referenzierte Methode höher (Tabelle 7-25). Ausnahmen zeigten sich in der körperlichen Belastung und der Leistung.

	screen-referenziert	world-referenziert	p-value (p-test)
Geistige Belastung	59.41 (27.44)	43.09 (26.54)	0.007
Körperliche Belastung	24.27 (17.84)	31.32 (24.26)	0.08
Zeitliche Belastung	57.35 (30.23)	45.15 (26.04)	0.03
Leistung	45.74 (26.06)	44.71 (26.97)	0.4
Anstrengung	48.82 (28.04)	36.32 (22.47)	0.02
Frustration	46.32 (29.42)	30.15 (22.21)	0.006
Gesamtbelastung	46.98 (20.23)	38.45 (15.48)	0.38

Tabelle 7-25: Werte für die Items des NASA-TLX (Mittelwert, Standardabweichung). Die signifikanten Unterschiede zwischen den beiden Methoden wurden durch den t-Test berechnet und fett markiert. In all diesen Fällen sind die Scores für die screen-referenzierte Methode höher.

Für eine nähere Betrachtung wurden die Daten der beiden Videos getrennt untersucht (Abbildung 7-101). Der t-Test zeigte für die meisten Items im *meeting*-Video keinen signifikanten Unterschied. Bezüglich der körperlichen Belastung gab es jedoch einen höheren Score für die world-referenzierte Methode ($p = 0.001$).

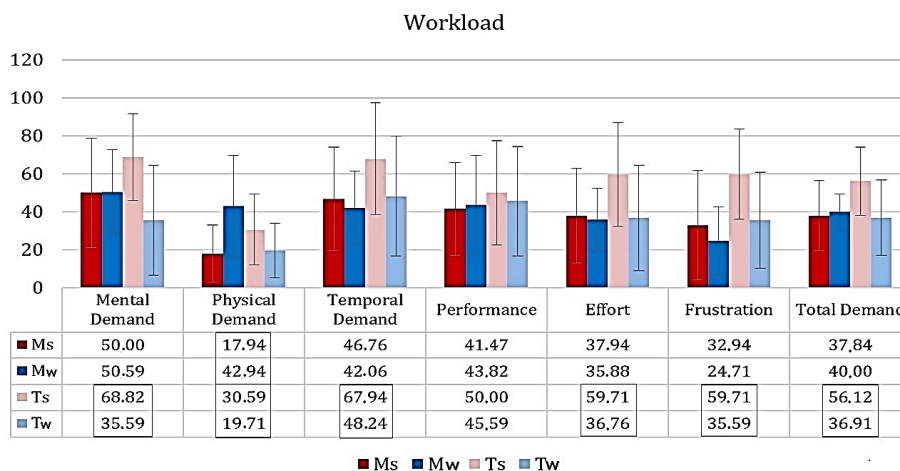


Abbildung 7-101: Ergebnisse des NASA-TLX-Fragebogens (Mittelwert, Standardabweichung) für jede Video-/Methodenkombination. Die signifikanten Unterschiede sind durch Umrandung gekennzeichnet.

Bei der Analyse der Daten des *talk*-Videos und der Durchführung von t-Tests für jedes Item wurde eine signifikant höhere Belastung für die screen-referenzierte Methode festgestellt:

- geistige Belastung ($p = 0.0004$)
- physische Belastung ($p = 0.032$)
- zeitliche Belastung ($p = 0.035$)
- Aufwand ($p = 0.01$)
- Frustration ($p = 0.004$)
- Gesamtbelastung ($p = 0.003$)

Die Unterschiede für alle anderen Positionen sind nicht signifikant.

Simulator-Sickness

Generell hatten die Teilnehmer nur leichte Beschwerden. Der Gesamtscore des SSQ war jedoch höher für die screen-referenzierte Methode. Betrachtet man die Ergebnisse bzgl. der Videos getrennt, ergeben sich

bei den meisten Items nur geringe Unterschiede, die Augenbelastung ist jedoch bei der screen-referenzierten Methode (Abbildung 7-102) signifikant höher (t-Test: $p = 0.047$).

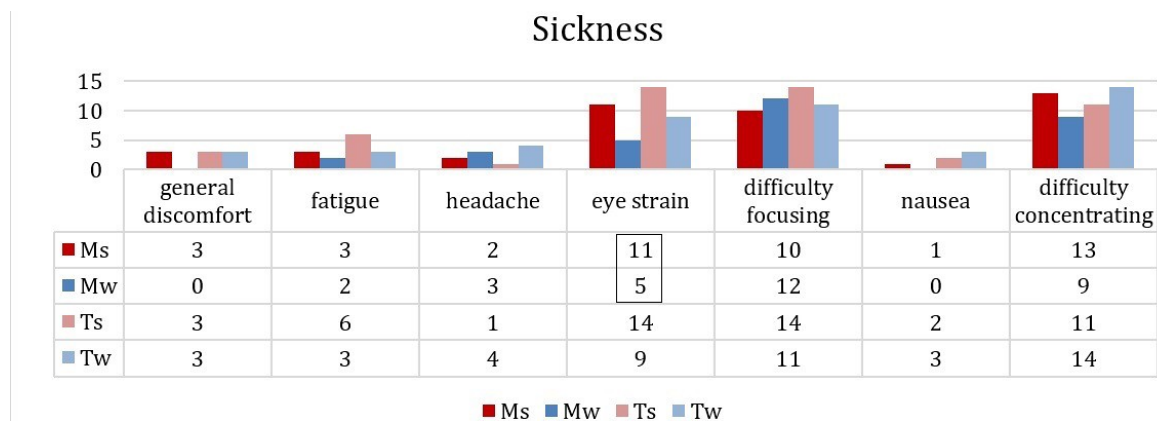


Abbildung 7-102: Scores (Summen) der Simulator-Sickness für jede Video- / Methodenkombination. Es gibt einen signifikanten Unterschied für die Augenbelastung.

Presence

Die Unterschiede in den Presence-Scores waren in fast allen Fällen gering. Der Score für die Frage „Wie schnell haben Sie sich an die virtuelle Umgebung gewöhnt?“ war jedoch für die world-referenzierte Methode signifikant höher (t-Test: $p = 0.01$) (Tabelle 7-26).

	screen-referenziert	world-referenziert	p(t-test)
How involved were you in the virtual environment experience?	3.88 (1.68)	4.06 (1.94)	0.3
How much did the visual aspects of the environment involve you?	4.24 (1.83)	4.18 (1.82)	0.4
How much did the auditory aspects of the environment involve you?	3.67 (1.72)	3.26 (1.68)	0.1
How quickly did you adjust to the virtual environment experience?	5.26(1.81)	6.15 (1.18)	0.01
How much did your experiences in the virtual environment seem consistent with your real-world experiences?	4.88 (1.70)	5.12 (1.61)	0.2

Tabelle 7-26: Mittelwerte und Standardabweichung für die Presence-Fragen. Signifikante Unterschiede zwischen den beiden Methoden sind fett markiert.

Um detailliertere Ergebnisse zu erhalten, wurden auch hier die Antworten für die beiden Videos getrennt analysiert (Abbildung 7-103) und für jedes Item ein t-Test durchgeführt. Wir fanden signifikant höhere Presence-Levels in den folgenden Kombinationen:

- Ts: Audio ($p = 0.041$)
- Mw: Involvement ($p = 0.011$)
- Mw: Zeit ($p = 0.014$).

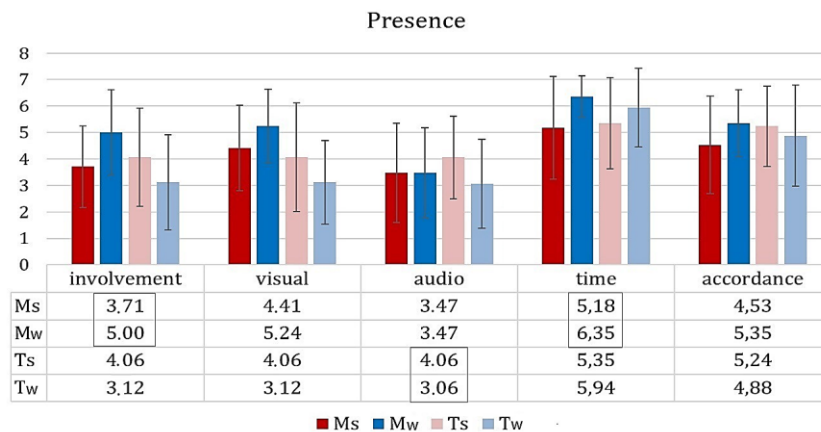


Abbildung 7-103: Mittelwerte und Standardabweichungen für die Presence-Fragen für jede Video- / Methodenkombination (Likert-Skala von 1 bis 7). Eingerahmte Werte zeigen signifikante Unterschiede zwischen den beiden Methoden.

In Tabelle 7-27 sind die Ergebnisse der Fragebogenanalyse zusammengefasst.

	talk	meeting	gesamt
Belastung (niedriger)	world	physisch: screen	world
Simulator-Sickness (niedriger)		Augenbelastung: world Allgemein: world	world
Presence (höher)	audio:screen	Involvierung: world Zeit: world	Zeit: world

Tabelle 7-27: Zusammenfassung der Ergebnisse. Die Tabelle zeigt die Methoden mit dem jeweils besten Ergebnis.

Qualitative Analyse

Nach jedem Abschnitt des Fragebogens konnten die Teilnehmer Kommentare zu den Methoden notieren. Zusätzlich wurde ein semistrukturiertes Interview durchgeführt. Es wurden die folgenden Aussagen gemacht.

Vorteile der screen-referenzierten Methode:

- „Ich kann entscheiden, wo ich hinschaue.“ (P10, P30, P11, P19, P23, P31, P34, P16, P32)
- „Diese Methode ähnelt der Methode im Fernsehen.“ (P18)
- „Die Untertitel sind immer sichtbar.“ (P5, P6, P2, P26)

Nachteile der screen-referenzierten Methode:

- „Es ist schwierig, die sprechende Person zuzuordnen.“ (P14)

Vorteile der world-referenzierten Methode:

- „Untertitel können dem Sprecher leichter zugewiesen werden.“ (P1, P25, P14, P3, P15, P8, P24)
- „Sprecher und Untertitel können gleichzeitig gesehen werden.“ (P33)
- „Es ist eine natürlichere Erfahrung.“ (P13, P15, P12, P24)
- „Es ist einfacher, dem Inhalt zu folgen.“ (P17, P21, P4, P8)

Nachteile der world-referenzierten Methode:

- „Ich bin gezwungen, auf den Sprecher zu schauen.“ (P7, P19)
- „Es ist manchmal schwierig, den Sprecher zu finden.“ (P29, P33)
- „Ich wusste nicht, wo der nächste Untertitel erscheinen wird.“ (P33)

Die direkte Frage nach den Präferenzen bezüglich der beiden Untertitelungsmethoden ergab ein ausgewogenes Ergebnis. Im Gegensatz zu den Ergebnissen aus den NASA-TLX-, SSQ- und Presence-Fragen, konnte keine bevorzugte Methode ermittelt werden. Wir haben jedoch wichtige Hinweise für Probleme erhalten, die in Zukunft untersucht werden sollten:

- Mit world-referenzierten Untertiteln ist es schwieriger sich umzuschauen.
- Mit world-referenzierten Untertiteln ist es schwieriger den nächsten Sprecher zu finden.
- Screen-referenzierte Untertiteln lassen sich schwieriger einem Sprecher zuordnen.

7.4.5.4 Einschränkungen

Die Ergebnisse dieser Studie sind von der Art der Szene beeinflusst. Es wurden zwei Arten von Szenen untersucht: ein Dialog zweier Personen und eine Gruppe von sprechenden Personen. Die Protagonisten haben während des Gesprächs ihre Position nicht geändert. Für sich bewegende, sprechende Protagonisten, müssen sich world-referenzierte Untertitel mitbewegen, was zu Übelkeit führen könnte. Solche Szenarien erfordern weitere Tests.

Bei der Analyse der Headtracking-Daten wurden für die Gesprächsszene Unterschiede zwischen den Methoden festgestellt. In Zeitintervallen, in denen Personen sprachen, konzentrierten sich die Blickrichtungen bei der world-referenzierte Methode mehr auf die Sprecher, was weniger Kopfbewegungen verursachte. Dies könnte ein Grund für die geringere Belastung sein, die aus den Antworten der NASA-TLX ersichtlich ist.

Die Teilnehmer dieser Studie waren hörende Menschen. Die Ergebnisse können daher hilfreich sein, um geeignete Untertitelmethoden für Fremdsprachen zu finden. Es sind weitere Experimente notwendig, um herauszufinden, ob gehörlose oder schwerhörige Zuschauer andere Ansätze benötigen (**Studie 7.4.7**).

Für die Erfassung der Blickrichtung wurde Headtracking verwendet. Die zusätzliche Verwendung eines Eyetrackers könnte zu detaillierteren Ergebnissen bei der Analyse der Blickrichtung führen.

Das Videomaterial unserer Nutzerstudie war sehr kurz (insgesamt 3 Minuten). Das längere Lesen von Untertiteln in CVR ist anstrengender und erfordert weitere Untersuchungen.

32 % der Teilnehmer waren VR-Anfänger. Das Lesen von Untertiteln in CVR war für alle neu. Eine Gewöhnung an CVR-Untertitel, kann zu geringerer Belastung und entspannterem Lesen führen.

7.4.5.5 Diskussion und Fazit

Beide Methoden - screen-referenzierte und world-referenzierte Untertitel - sind hilfreich, um Filme in Fremdsprachen zu verstehen. Auch wenn unsere Arbeit nur ein erster Ansatz ist und nur zwei spezielle Szenen untersucht wurden, ermutigen die Ergebnisse zu weiteren Studien auf diesem Gebiet. Wir glauben, dass world-referenzierte Untertitel, die derzeit in CVR nicht verwendet werden, viel Potenzial haben. Keine der untersuchten Methoden erfüllt jedoch alle Anforderungen für jedes Szenario in CVR. Eine Kombination der Methoden könnte ein verbesserter Ansatz sein (**Studie 7.4.8**). Ferner könnten die Untertitelungsverfahren durch Techniken der Aufmerksamkeitslenkung erweitert werden, um die Sprecheridentifikation zu erleichtern.

Darüber hinaus sollten die Untersuchungen mit gehörlosen oder schwerhörigen Menschen fortgesetzt werden, was weitere Überlegungen zur Identifizierung der Sprecher erfordert. Für hörende Menschen sind die Stimmen der Protagonisten dafür eine Hilfe, die gehörlosen Menschen nicht zur Verfügung steht. Verschiedene Farben, Schriftarten oder Zeichen werden bereits für die Untertitelung konventioneller Filme genutzt und können angepasst werden. Das Problem der Sprecheridentifikation in CVR ist jedoch schwieriger als in herkömmlichen Filmen und bedarf weiterer Untersuchungen (**Studien 7.4.6 und 7.4.7**).

7.4.6 Studie: Sprecheridentifikation für Gebärdensprache

Die Studie ist Teil der Bachelorarbeit von Lili Tran (2017): Embedding of Sign Language in Cinematic VR.

Ich habe diese Arbeit betreut und gemeinsam mit der Studentin die Idee entwickelt, die sie implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir von der Studentin durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Das Ziel dieser Studie war es, Methoden zu finden, die Menschen mit eingeschränktem Hörvermögen dabei unterstützen, den Gebärdendolmetscher der sprechenden Person im Film zuzuordnen. Dafür wurden zwei Methoden implementiert und miteinander verglichen.

border

Für die *border*-Methode wird das Bild des Gebärdendolmetschers farblich umrandet. Jeder sprechenden Person wird dabei eine andere Farbe zugeordnet (Abbildung 7-104, links)

image

Für die *image*-Methode wird dem Gebärdendolmetscher ein Bild des Sprechers beigelegt (Abbildung 7-104, rechts).



Abbildung 7-104: links: Die *border*-Methode kennzeichnet den Sprecher mit einer farbigen Umrandung. Rechts: Die *image*-Methode umfaßt ein Bild des Sprechers.

7.4.6.1 Material

Für die Tests wurden zwei Animationsvideos ausgewählt, in denen verschiedene Personen sprechen, die nicht gleichzeitig im FoV sichtbar sind. In einem der beiden Videos gab es zwei sprechende Charaktere (Walt Disney Studios, 2016), im anderen fünf (Lego, 2017). Beide Videos hatten eine ungefähre Länge von 2 min. Die Teilnehmer sahen die Videos über eine HTC Vive. Für die Testumgebung und die Implementierung der Methoden wurde Unity 3D verwendet.

7.4.6.2 Parameter - Vorstudie

In einer informellen Vorstudie mit drei Teilnehmern wurden die Parameter für die Methoden getestet. Für die *border*-Methode wurde das Fenster der *PiP*-Methode aus **Studie 7.1.3** übernommen. Um die zentralen Bildinhalte, die der Nutzer in der Regel in der Mitte des Displays platziert, nicht zu überdecken wurde das Dolmetscherfenster zunächst am unteren rechten Rand platziert. Dies erwies sich allerdings

als ungünstig, da diese Stelle in der HTC Vive unschärfer ist als mittige Bereiche. Außerdem konnte dieser Bereich von den Teilnehmern schlecht fokussiert werden. Das Fenster wurde deshalb weiter oben platziert. Die Fenstergröße, die von den Teilnehmern als zu groß eingeschätzt wurde, konnte nicht verkleinert werden, da dann Details der Gebärdensprache nicht mehr ausreichend erkennbar waren.

7.4.6.3 Teilnehmer und Prozedur

Sechs Personen (4 weiblich, 2 männlich, Alter: MW=47.8 Jahre) nahmen an der Studie teil. Vier der Teilnehmer waren gehörlos und zwei schwerhörig. Für fünf Teilnehmer war Gebärdensprache der normale Kommunikationsweg. Die Teilnehmer hatten keine oder wenig VR-Erfahrungen.

Ein within-subject Testdesign wurde angewendet. Jeder Teilnehmer sah beide Videos, jedes mit einer anderen Methode. Die Reihenfolge der Videos und Methoden war ausbalanciert. Die folgenden abhängigen Variablen wurden für jede der beiden Methoden per Fragebogen ermittelt: Simulator-Sickness, Presence und Usability. Der Fragebogen endete mit Fragen zum Vergleich der beiden Methoden. Es wurden die folgenden Usability-Fragen gestellt:

- War es dir möglich dem Dolmetscher und dem Videoinhalt zu folgen?
- Wich deine Aufmerksamkeit wegen des Dolmetschers vom Video ab?
- Findest du Positionierung und Größe des Dolmetschers passend?
- Konntest du die Texte des Dolmetschers den Sprechern zuordnen?

Der Fragebogen endete mit Fragen zum Vergleich der beiden Methoden.

7.4.6.4 Ergebnisse

Simulator-Sickness, Presence

Es konnten keine signifikanten Unterschiede zwischen den beiden Methoden festgestellt werden.

Usability

Die Ergebnisse waren für beide Methoden ähnlich. Fünf Teilnehmer konnten dem Inhalt des Videos nicht folgen. Alle Teilnehmer gaben an, dass der Gebärdendolmetscher vom Video ablenkte. Die Kommentare der Teilnehmer können dabei helfen, beide Methoden zu verbessern:

- Ungeeignete Hintergrundfarbe des Gebärdendolmetschers (P1)
- Gebärdendolmetscher verdeckt zu viel (P2)
- Gebärdendolmetscher besser ohne Hintergrund (P5, P6)
- Gebärdendolmetscher unscharf (P3, P5)

Vergleich

Vier Teilnehmer fanden die *image*-Methode hilfreicher, für zwei Teilnehmer waren beide Methoden gleich.

7.4.6.5 Diskussion

Die *image*-Methode war für die meisten Teilnehmer hilfreicher als die *border*-Methode. Da die Randbereiche auf HTC-Vive-Displays momentan nur unscharfe Darstellungen ermöglichen ist es schwer, eine geeignete Position des Gebärdendolmetschers zu finden, die den Inhalt des Films nicht wesentlich überdeckt.

Alle Teilnehmer hatten Schwierigkeiten, dem Gebärdendolmetscher und dem Video gleichzeitig zu folgen. Weiter Untersuchungen sind notwendig, um Methoden zu finden, wie Gebärdendolmetscher in CVR mit mehreren Sprechern eingebettet werden können.

7.4.7 Studie: Sprecheridentifikation für Untertitel

Die Studie ist Teil der Bachelorarbeit von Lisa Görtz, (2017): Methods of Speaker Identification for Subtitles in Cinematic VR.

Ich habe diese Arbeit betreut und gemeinsam mit der Studentin die Idee entwickelt, die sie implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir von der Studentin durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Das Ziel dieser Studie war es, Methoden zu finden, die Menschen mit eingeschränktem Hörvermögen dabei unterstützen, eine CVR-Applikation mit Untertiteln zu verfolgen und dabei die Untertitel der sprechenden Person im Film zuzuordnen. Die Studie verwendete ähnliche Methoden wie **Studie 7.4.5** mit dem Unterschied, dass statt den Namen Farben zur Sprecherzuordnung der Untertitel verwendet wurden. Es wurden zwei Methoden implementiert und miteinander verglichen.

screen-referenzierte Untertitel

Für die screen-referenzierte Methode werden die Untertitel am Display fixiert und drehen sich mit der Blickrichtung des Zuschauers mit.

world-referenzierte Untertitel

Für die world-referenzierte Methode werden die Untertitel in der Nähe des jeweiligen Sprechers positioniert und sind in der CVR-Welt verankert.

7.4.7.1 Material

Es wurden die gleichen Videos wie in **Studie 7.4.6** genutzt: zwei Animationsvideos, in denen verschiedene Personen sprechen, die nicht gleichzeitig im FoV sichtbar sind. In einem der beiden Videos gab es zwei sprechende Charaktere (Walt Disney Studios, 2016), im anderen fünf (Lego, 2017). Beide Videos hatten eine ungefähre Länge von 2 min. Die Teilnehmer sahen die Videos über eine HTC Vive. Für die Testumgebung und die Implementierung der Methoden wurde Unity 3D verwendet.

7.4.7.2 Teilnehmer und Prozedur

Sechs Personen (4 weiblich, 2 männlich, Alter: MW=47.8 Jahre) nahmen an der Studie teil. Vier der Teilnehmer waren gehörlos und zwei schwerhörig. Für fünf Teilnehmer war Gebärdensprachen der normale Kommunikationsweg. Die Teilnehmer hatten keine oder wenig VR-Erfahrungen. Es wurde ein within-subject Testdesign angewendet und die folgenden abhängigen Variablen für jede der beiden Methoden per Fragebogen ermittelt: Simulator-Sickness, Presence und Usability. Es wurden die folgenden Fragen zur Usability gestellt:

- Fandest du die Untertitel ansprechend zum Anschauen?
- Hattest du Probleme Untertitel und Videoinhalt gleichzeitig zu verfolgen?
- Konntest du den jeweiligen Sprecher immer identifizieren?
- Kannst du dir vorstellen, einen längeren Film mit diesen Untertiteln anzuschauen?

Der Fragebogen endete mit Fragen zum Vergleich der beiden Methoden.

7.4.7.3 Ergebnisse

Simulator-Sickness, Presence

Es konnten keine signifikanten Unterschiede zwischen den beiden Methoden festgestellt werden.

Usability

Für die meisten Items unterscheiden sich die Antworten zwischen den beiden Methoden nicht. Allerdings gaben mehr der Teilnehmer bei der world-referenzierte Methode an, Schwierigkeiten bei der Identifizierung der Sprecher zu haben, als bei der screen-referenzierten.

Vergleich

Alle Teilnehmer würden sich mit beiden Methoden auch längere Filme anschauen. Für vier Teilnehmer war die Sprecheridentifizierung mit screen-referenzierten Untertiteln leichter, für zwei mit world-referenzierten. Alle Teilnehmer fanden, dass die screen-referenzierten Untertitel leichter zu lesen sind. Zur Verbesserung der Sprecheridentifikation wurden für die screen-referenzierten Untertitel zusätzliche Pfeile empfohlen, die in die Richtung der Sprecher zeigen. Vier Teilnehmer wünschten sich einen Wechsel zwischen den beiden Methoden, den wir in **Studie 7.4.8** untersuchen.

7.4.7.4 Diskussion

Beide Methoden wurden von den Teilnehmern akzeptiert. Die Ergebnisse hängen in großem Maße von den gewählten Parametern wie Schriftgröße und Platzierungen ab, für die weitere Untersuchungen und Anpassungen notwendig sind. Auch die gewählten Filmszenarien beeinflussen die Ergebnisse. Eine größere Anzahl von Personen erschwert die Sprecheridentifikation und benötigt neue Herangehensweisen. Die screen-referenzierten Untertitel waren für die Teilnehmer leichter zu handhaben, was einerseits daran liegen kann, dass diese dem bisherigen Sehverhalten entsprechen. Andererseits scheinen world-referenzierte Untertitel für gehörlose Personen schwieriger zu handhaben sein, als für hörende. Ist der nächste Sprecher nicht im FoV, bemerkt der Zuschauer in diesem Fall nicht, dass eine andere Person spricht. Die Mehrzahl der Zuschauer schlug Pfeile in Sprecherrichtung als zusätzliche Hilfe vor.

7.4.8 Studie: Kombination world- und screen-referenzierter Untertitel

Die Studie ist Teil der Bachelorarbeit von Michaela Schmierl (2018): Enhanced Subtitle Methods in Cinematic Virtual Reality.

Ich habe diese Arbeit betreut und gemeinsam mit der Studentin die Idee entwickelt, die sie implementierte und testete. Die Projektschritte wurden in regelmäßigen Treffen besprochen und die Studie in Abstimmung mit mir von der Studentin durchgeführt. Die Daten habe ich neu analysiert und interpretiert.

Das Ziel dieser Studie war es, Methoden zu finden, die Menschen mit eingeschränktem Hörvermögen dabei unterstützen, eine CVR-Applikation mit Untertiteln zu verfolgen und dabei die Untertitel der sprechenden Person im Film zuzuordnen. Dafür wurden die beiden Methoden aus **Studie 7.4.5** miteinander kombiniert, so dass zwischen ihnen dynamisch gewechselt wurde. Die world-referenzierten Untertitel waren wie in **7.4.5** und **7.4.7** neben dem Sprecher platziert. Schaute der Betrachter in eine andere Richtung, so dass die Untertitel nicht im FoV waren, wurde automatisch auf screen-referenzierte Untertitel umgeschaltet. Wie in **Studie 7.4.7** wurden zur Unterstützung der Sprecheridentifikation unterschiedliche Farben für die Personen verwendet. Durch den Farbwechsel kann auch eine gehörlose Person den Sprecherwechsel erkennen.



Abbildung 7-105: Zusätzliche Cues weisen auf die sprechende Person. links: *line*-Methode, rechts: *arrow*-Methode.

Wie in **Studie 7.4.7** empfohlen, werden für die Untertitel zusätzliche Methoden eingeführt, um auf die sprechende Person hinzuweisen. Dabei wurde zum einen die *line*-Methode aus **Studie 7.3.4** genutzt und mit einer neu designten *arrow*-Methode verglichen (Abbildung 7-105).

line

Es wird eine Linie von den Untertiteln zur sprechenden Person angezeigt.

arrow

Es wird ein Pfeil, der neben die Untertitel platziert ist, angezeigt, der in Richtung der sprechenden Person weist.

7.4.8.1 Material

Es wurden die gleichen Videos wie in **Studie 7.2.2** genutzt: drei Teile eines Animationsvideos, in denen zwei Personen miteinander sprechen, die nicht gleichzeitig im FoV sichtbar sind (*wide-Seitenansicht*). Das Video hat eine Gesamtlänge von 4 min. Die Teilnehmer sahen die Videos über ein Fove-HMD mit integriertem Eyetracker. Für die Implementierung der Methoden ist Unity 3D verwendet worden. Der Ton der Videos war stumm geschaltet, um die Teilnehmer zu animieren, die Untertitel zu lesen.

7.4.8.2 Parameter

In einer informellen Vorstudie mit sieben Teilnehmern testeten wir verschiedene Parameter getestet, unter anderem verschiedene Pfeilarten. Die Teilnehmer bevorzugten einen 2D Pfeil, der im Inneren durchsichtig ist und dadurch weniger Bildinhalt verdeckt. Sowohl Pfeil als auch Schriftgröße wurden als Ergebnis der Vorstudie verkleinert und der Pfeil näher an den Untertiteln platziert.

7.4.8.3 Teilnehmer und Prozedur

Die Studie wurde mit Teilnehmern mit und ohne Höreinschränkungen durchgeführt. Die Daten haben wir getrennt analysiert.

Am ersten Teil der Studie nahmen 34 hörende Personen teil (21 weiblich, 13 männlich, Alter: MW= 35.5, SD=18.9). 66.7 % der Teilnehmer hatten keine oder wenig VR-Erfahrungen.

Ein within-subject Testdesign wurde angewendet und die Reihenfolge der Filme und Methoden ausbalanciert. Mittels Fragebogen ermittelten wir die folgenden abhängigen Variablen für jede der beiden Methoden: Presence, Simulator-Sickness und Usability. Der Fragebogen endete mit abschließenden Fragen zu beiden Methoden.

Am zweiten Teil der Studie nahmen zwei Personen mit Höreinschränkungen des Berufsbildungswerks München teil: eine 22-jährige männliche Person und eine 20-jährige weibliche Person. Dafür wurden Ergebnisse des ersten Teils der Studie berücksichtigt und die Fragen leicht modifiziert.

7.4.8.4 Ergebnisse

Simulator-Sickness, Presence

Es konnten keine signifikanten Unterschiede zwischen den beiden Methoden festgestellt werden.

Usability

Für zwei Items stellten wir signifikante Unterschiede in den Scores fest (Abbildung 7-106, Tabelle 7-28). In der Baseline hatten mehr Teilnehmer das Problem, den Sprecher zu identifizieren als in den beiden Bedingungen, in denen Guiding Methoden benutzt wurden (arrow/line).

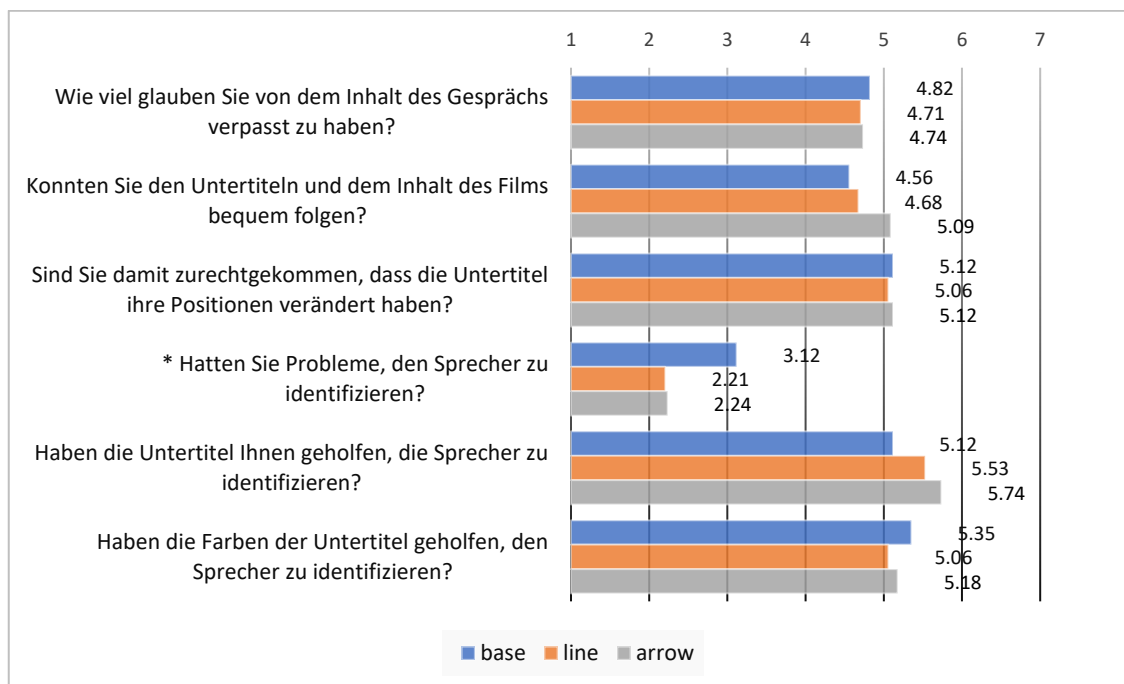


Abbildung 7-106: Mittelwerte der Experience-Items (Likert-Skala von 1 bis 7). Items mit signifikanten Unterschieden sind mit * gekennzeichnet.

	arrow	line	base	p-value
Hatten Sie Probleme, den Sprecher zu identifizieren?	2.24 (1.5)	2.21 (1.53)	3.12 (1.99)	$p_{\text{base/line}}=0.05$ $p_{\text{base/arrow}}=0.03$

Tabelle 7-28: Mittelwerte, Standardabweichungen (in Klammern) und p-Werte für die Fragen mit signifikanten Unterschieden zwischen den Methoden.

Abschließende Fragen

Teil 1 (hörende Teilnehmer): Die Mehrheit der Teilnehmer würde sich mit beiden Methoden auch längere Filme anschauen. Die Linie empfanden die Teilnehmer ablenkender ($MW_{\text{line}} = 6.03$, $SD_{\text{line}} = 1.3$) als den Pfeil ($MW_{\text{arrow}} = 2.47$, $SD_{\text{arrow}} = 1.6$, $p < 0.001$).

Die meisten Teilnehmer (79.41 %) würden es nicht vorziehen, dass screen-referenzierten Untertitel auch angezeigt werden, wenn die world-referenzierten Untertitel sichtbar sind. 55.8 % der Teilnehmer empfanden den Pfeil hilfreicher, um zum Sprecher gelenkt zu werden. Teil 2 (Teilnehmer mit Höreinschränkungen): Die Ergebnisse sind ähnlich zu denen des ersten Teils. Auch die Teilnehmer mit eingeschränktem Hörvermögen fühlten sich durch den Pfeil weniger abgelenkt und möchten nicht, dass die screen-referenzierten Untertitel gemeinsam mit den world-referenzierten Untertiteln sichtbar sind.

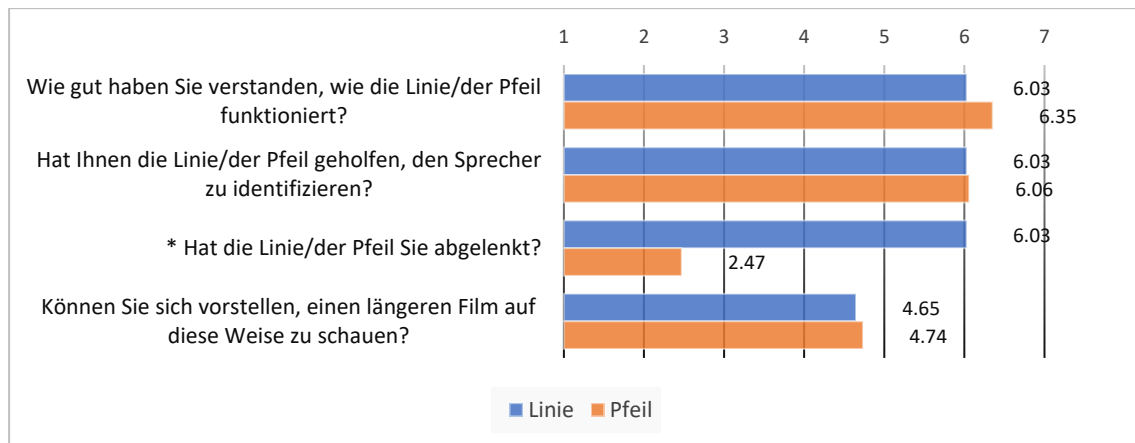


Abbildung 7-107: Mittelwerte für die abschließenden Fragen (Likert-Skala von 1 bis 7). Signifikante Unterschiede sind mit * gekennzeichnet.

7.4.8.5 Diskussion

Durch die verwendeten Methoden konnte die Sprecheridentifikation in Bezug auf die Baseline verbessert werden. Der automatische Wechsel zwischen screen- und world-referenzierten Untertiteln funktionierte gut, wobei sich der Pfeil als hilfreicher und weniger ablenkend erwies als die Linie, um zum Sprecher gelenkt zu werden.

7.5 Fazit

Die vorgestellten Nutzerstudien sind ein Schritt, mehr Erfahrungen über das Medium Cinematic Virtual Reality zu sammeln. Es wurde deshalb versucht die durchgeführten Experimente möglichst vollständig wiederzugeben, auch wenn nicht jede dieser Studien neue effektive Methoden hervorbrachte. Aber auch die Erfahrungen, dass manche Ideen nur wenig oder keine Fortschritte bewirkten, sind Erkenntnisse, die notwendig sind und die Entwicklung des Mediums voranbringen. Gründe zu suchen, warum eine Methode nicht bei jedem Menschen so funktionierte, wie erwünscht, halfen uns weiter und vermittelten uns neue Erkenntnisse, um immer wieder neue Dinge auszuprobieren. Die Studien zum Nutzerkomfort lieferten wichtige Erkenntnisse, wie CVR-Applikationen für den Zuschauer angenehmer gestaltet werden können, und viele der vorgestellten Techniken zur Aufmerksamkeitslenkung, Kamerapositionierung und Ausrichtung von Schnitten sind Mittel, um die Filmsprache für CVR weiterzuentwickeln.

Schlusswort

In der vorliegenden Arbeit wurde das Thema Cinematic Virtual Reality von unterschiedlichen Seiten betrachtet. Es wurden sowohl Aspekte des Produktionsprozesses, des Nutzerkomforts wie auch der Datenanalyse berücksichtigt. Das vorgestellte Spaceline-Konzept kann mit dazu beitragen eine neue Filmsprache für CVR zu entwickeln und dabei die individuellen Anforderungen und unterschiedlichen Sehverhalten der Zuschauer beachten. Es nutzt die neuen Möglichkeiten, die CVR bietet und verknüpft sie mit bewährten Techniken der traditionellen Filmproduktion. Durch die interaktive Komponente kann jeder Zuschauer seinen eigenen Film erleben, in seinem eigenen Tempo und mit eigenen Schwerpunkten.

Um dieses Konzept umzusetzen, braucht es Tools zum Erstellen der Applikationen und Player, die die Interaktionen der Zuschauer berücksichtigen. Für beides wurden Konzepte vorgestellt.

CVR ist noch ein recht neues Forschungsgebiet ohne lang erprobte Forschungsmethoden und Grundlagenkenntnisse und es entwickelt sich wahnsinnig schnell. Mehrere Male machten wir die Erfahrung, dass andere gleichzeitig an ähnlichen Problemen arbeiteten und konnten unsere gesammelten Erkenntnisse mit anderen Wissenschaftlern austauschen. Die wissenschaftliche CVR-Community ist in den letzten Jahren gewachsen und kann viel von dem Wissen aus Filmtheorie, VR und AR lernen, um dieses neue Forschungsgebiet weiterzuentwickeln. Aber auch umgekehrt, kann das in CVR erworbene Wissen in anderen Gebieten angewendet werden. Unsere Taxonomie für Guiding-Methoden kann zum Großteil auch für AR- und VR-Anwendungen verwendet werden, die Guidelines für Kamerapositionen sind auch auf virtuelle Kameras in VR-Umgebungen übertragbar und auch das Konzept gerichteter Links kann für VR-Transitions angewendet werden.

Immer wieder konnten wir in unseren Studien feststellen, dass Teilnehmer, von diesem neuen Medium fasziniert sind. Sicher hat der Novelty-Effekt einige unserer Testergebnisse beeinflusst. Aber es ist auch etwas Besonderes, die Entstehung einer neuen Ausdrucksform von Anfang an erforschen zu können, mit unvoreingenommenen Zuschauern, deren Sehgewohnheiten noch nicht von diesem Medium geprägt wurden. Es wird auch weiterhin spannend, die Entwicklung von CVR wissenschaftlich zu begleiten und neue Wege und Ausdrucksweisen zu finden.

Anhang

Begriffe

Da ein Teil der Terminologie in verschiedenen Fachgebieten unterschiedlich genutzt wird, wird in dieser Übersicht zusammengefasst, wie die Begriffe in der vorliegenden Arbeit verwendet wurden. Begriffe, die in dieser Arbeit neu eingeführt wurden, sind *kursiv* dargestellt.

<i>add-Region</i>	Element des Spaceline-Konzeptes: Region eines Spaces, bei deren Aktivierung zusätzliche Text-, Audio-, Bild- oder Video-Informationen angezeigt werden
äquirektanguläre Projektion	Projektion der Kugeloberfläche auf die Ebene, bei der die Winkelkoordinaten (Länge, Breite) direkt als kartesische Koordinaten verwendet werden. Sie wird oft genutzt, um die Weltkugel als Karte darzustellen und deshalb auch geographische Projektion genannt.
<i>apparent</i>	offensichtlich, deutlich wahrnehmbar. Apparente Methoden sind im Gegensatz zu subtilen Methoden vom Nutzer wahrnehmbar und der Nutzer ist sich der Methode bewusst.
AR- Augmented Reality	Erweiterte Realität: Überlagern einer virtuellen Ebene über die Ansicht der realen Welt, Mischung aus Realität und computergenerierten Inhalten als eine Erweiterung der realen Welt mit virtuellen Informationen
Awareness	sich etwas bewusst sein
Bildausschnitt	Teil des Motivs, welches der Zuschauer sieht
Cardboard	Halterung für Smartphones aus Karton oder Kunststoff für 3DoF Virtual Reality Anwendungen
CAVE	cave automatic virtual environment, immersive, virtuelle Realitätsumgebung, in der Projektoren auf Wände eines Raumes VR-Inhalte projizieren
Cinematic VR	Erleben von sphärischen Filmen mit HMDs oder anderen VR-Geräten. Ein Ziel von CVR ist es, den Zuschauer an einen virtuellen Ort zu transportieren und ihm das Gefühl zu geben, innerhalb der Szene zu sein.
co-located	am gleichen Ort
Co-Watcher	Person, mit der gemeinsam ein Film angeschaut wird
Container-Format	Dateiformat mit festgelegten Datenstrukturen, welches den Austausch von Daten zwischen Programmen ermöglicht. Ein Container enthält verschiedene Dateien unterschiedlicher Formate und Metadaten.
Cue	Hinweis, Markierung, Zeichen
CVR	Cinematic Virtual Reality, Cinematic VR
CVR-Entwickler	Person, die an der Schaffung von CVR-Applikationen beteiligt ist. Dies umfasst unter anderem: Autoren, Kameralleute, CVR-Designer, CVR-Programmierer.
deskriptive Statistik	beschreibende Statistik, Beschreibung von Testergebnissen durch Kennwerte (z.B. Mittelwerte), Tabellen und Grafiken

Design-Space	Tool, welches die relevanten Dimensionen zum Entwerfen eines Produktes aufzeigt und damit den Designprozess unterstützt und erweitert. Ein Design Space reflektiert die Eigenschaften und die Designentscheidungen vorhandener Produkte, die für einen ähnlichen Zweck verwendet werden. Er unterstützt die Erstellung neuer Produkte entlang einer Reihe von Dimensionen, für die mehrere Werte vorgeschlagen werden. Neue Designerfahrung werden umgesetzt, indem die vorhandenen Dimensionen mit neuen Designmöglichkeiten erweitert werden.
<i>dis-located</i>	an verschiedenen Orten
diegetisch	Diegetische Elemente sind Teil der Story-Welt, z. B. eine Musik aus einem Radio im Film. Eine Filmmusik ist dagegen nicht diegetisch.
DoF	Anzahl der voneinander unabhängigen Bewegungsdimensionen in einer VR-Welt
Dwellzeit	Zeitfenster bevor ein vordefiniertes Ereignis eintritt
Editing	Bearbeiten und Zusammenführen digitaler Daten zu einem neuen Werk oder Produkt
Einstellung	Folge von Einzelbildern, die ohne Unterbrechung aufgenommen wurden framed Filme: Filmsegment zwischen zwei Schnitten sphärische Filme: Bildsequenz in einem Space, die der Betrachter im FoV wahrnimmt
Einstellungsgröße	Größenverhältnis des abgebildeten Subjekts/Objekts zum vorgegebenen Bildfeld. Die Einstellungsgröße ergibt sich aus der Distanz der Kamera zum aufgenommenen Subjekt/Objekt und den gewählten Abbildungsparametern der Kamera.
Enjoyment	Zustand oder Prozess des Vergnügens, z.B. an einer digitalen Applikation
Experience	Erlebnis, welches ein Benutzer durch Interaktion über eine digitale Technologie mit einer Applikation hat
Eye-Tracker	Gerät zum Aufzeichnen von Augenbewegungen
Flicker	schnelle Folge von optischen Reizen mit wechselnden Eigenschaften, z. B. hell/dunkel-Flicker, rot/blau-Flicker
forciert	erzungen, nicht auf freien Willen basierend
FoV Field of View	Sichtfeld: Bereich, den eine Person ohne Kopfbewegung mit den Augen oder durch ein optisches Gerät sehen kann
<i>framed Film</i>	traditionelle Filme auf Leinwand, TV-Gerät oder Monitor. Die Bildsprache dieser Filme beruht unter anderem darauf, dass dem Zuschauer bestimmte Bildausschnitte (innerhalb eines Rahmens) präsentiert werden.
Degrees of Freedom (DoF)	Freiheitsgrade in VR, CVR hat im allgemeinen 3DOF, VR-Umgebungen, in den man sich fortbewegen kann 6DoF
<i>gerichtete Transitions</i>	Element des Spaceline-Konzeptes: Am Beginn einer neuen Einstellung schaut der Zuschauer in eine vom Filmemacher im Voraus bestimmte Richtung.
Guiding	Lenkung von Blick oder Aufmerksamkeit
harter Schnitt	Bildwechsel ohne Effekt-Übergang
Head-Tracker	Gerät zum Aufzeichnen von Kopfbewegungen

Heatmap	Visualisierung von Daten, bei der Häufigkeiten durch Farben dargestellt werden. Im Fall von Eyetracking-Daten bedeutet eine rote Fläche z. B., dass dieses Gebiet besonders viel angeschaut wurde.
HMD	Head-Mounted Display, am Kopf montierte Displays wie z. B. Virtual-Reality-Headsets
Immersion	Eintauchen in eine virtuelle Welt, Ziel ist es ein Gefühl der Präsenz erzeugen, wird beeinflusst durch die technischen Parameter eines VR-Systems
Indikator	Element des Spaceline-Konzeptes: Anzeiger, um auf Rols oder Regionen hinzuweisen
Inferenzstatistik	induktive Statistik. Dabei werden statistische Testverfahren verwendet, um aus den Befunden einer Stichproben auf Grundgesamtheiten zu schließen.
in-Region	Element des Spaceline-Konzeptes: Region, die der Zuschauer zu Beginn einer Einstellung sieht
Insert	Element des Spaceline-Konzeptes: zusätzliche Bild, Audio oder Text-Einblendung in einem Film oder Medienanwendung, die zusätzliche Informationen liefert
in-situ	Projektion von Elementen in die reale Welt mittels eines AR-Displays (Gegensatz zu in-view)
Interaktion	Vorgang, bei dem Personen oder Dinge miteinander kommunizieren oder aufeinander reagieren HCI: Vorgang, bei dem Personen und Computer miteinander kommunizieren und aufeinander reagieren
in-view	Projektion von Elementen auf ein AR-Display ohne räumliche Verbindung zur realen Welt (Gegensatz zu in-view)
Kadrierung	Festlegung des gezeigten Bildausschnittes
Link	Verknüpfung digitaler Daten; Element des Spaceline-Konzeptes: Verbindung zwischen zwei Spaces
Media-Editor	Programm zum Bearbeiten und Zusammenfügen von Mediendaten (z. B. Video, Audio)
Media-Player	Programm zum Abspielen von Mediendaten (z. B. Video, Audio)
Methode	Informatik: Prozeduren / Funktionen, die das Verhalten von Objekten beschreiben Psychologie: wissenschaftliches Vorgehen, um neue Erkenntnisse zu gewinnen
Mise-en-scène	Bildkomposition und Bildgestaltung im sichtbaren Bildausschnitt. Dies umfasst u. a. die Farbkomposition, die Lichtgestaltung, die Anordnung der Figuren und Dinge im Bild, die Ausstattung und die Schauspielerführung.
Montage	kreativer Prozess des Aneinanderfügens von Bildern und Tönen im Aufbau des Films
Novelty-Effekt	Wirkung einer Applikation bei Einführung einer neuen Technologie, die auf das höhere Interesse an der neuen Technologie zurückzuführen ist und mit zunehmender Vertrautheit abnimmt
off-screen	Element des Spaceline-Konzeptes: außerhalb des momentanen Sichtfeldes
omnidirektional	in jede Richtung
on-screen	Element des Spaceline-Konzeptes: im momentanen Sichtfeld

out-Region	Element des Spaceline-Konzeptes: Region, deren Aktivierung eine Einstellung
PiP	Bild-in-Bild (engl.: Picture-in-Picture): spezieller Mehrfenstermodus, der besonders bei Videowiedergaben verwendet wird. Dabei ist ein kleines Video-Fenster auf dem größeren Fenster des anderen Videos eingeblendet.
pitch / Nicken	Drehung um die y-Achse eines Referenzsystems
Pointing	Positionsänderung eines digitalen Zeigers durch Bewegung von Anzeigegegeräten oder Tracking (z. B. Computermouse, Kopf, Augen, Hände)
Presence	subjektives Gefühl, in einer virtuellen Umgebung „vor Ort“ zu sein psychologische, wahrnehmungsbezogene Folge der Immersion
Proxemik	wissenschaftliches Fachgebiet, welches sich mit dem Verhalten von Menschen im Raum als nonverbalen Kommunikationsform beschäftigt
redirect Walking	VR-Fortbewegungstechnik, mit der Benutzer eine virtuelle Welt erkunden können, die größer ist als der getrackte Arbeitsbereich. Der Benutzer wird durch Manipulationen der Szene umgeleitet, was dazu führt, dass Benutzer die Szenenbewegung unwissentlich kompensieren, indem sie sich neu ausrichten.
remote	ohne direkten physikalischen Zugriff, mittels Computernetzwerk miteinander verbunden
Roi	Region of Interest: Bereich eines (Film)-Bildes, der von Interesse ist
Salincy	Auffälligkeit
Saliency Map	zeigt die Saliency-Werte (Auffälligkeiten) für jede Region in einem Bild an
Saliency-Modulations-technik (SMT)	Methode der Aufmerksamkeitslenkung, bei der eine Region so moduliert wird, dass sie durch bestimmte Merkmale (z. B. Schärfe, Farbe, Helligkeit) aus der Umgebung hervorsticht
salient	hervorstechend, auffallend
Schnitt	(1) Übergang zwischen zwei Einstellungen (2) Phase der Filmproduktion, in der aus dem gedrehten Material der Film zusammengefügt wird
Screen	Bildschirm, Display
screen-referenziert	mit dem Display verbunden. Screen-referenzierte Objekte bewegen sich mit, falls der Betrachter den Kopf dreht
Sensor	technisches Bauteil zur Bestimmung spezieller Umgebungsmerkmale
sensorisch	Wahrnehmung von Reizen über die Sinnesorgane
Shape	Form einer digital erzeugten Fläche
Simulator-Sickness	Form von Unwohlsein, die bei VR-Applikationen auftreten kann, ähnelt Reisekrankheit
Soziale Presence	Gefühl des Zusammenseins in einer virtuellen Welt
Social Viewing	Szenarium, in dem mehrere Nutzer (Co-Watcher) gleichzeitig einen Film sehen und dabei die Möglichkeit haben miteinander zu kommunizieren
Space	Element des Spaceline-Konzeptes: sphärisches Filmsegment, das ohne Unterbrechung aufgezeichnet wurde

Spaceline	vom Betrachter gewählter Pfad durch eine Struktur von Spaces, welche vom Filmemacher entworfen wurde. Im Gegensatz zu einem timeline-basierten Film, der allein vom Filmemacher bestimmt wird, wird die Spaceline vom Filmemacher und vom Betrachter bestimmt.
spacelinebasierte Schnitte	Element des Spaceline-Konzeptes: Schnitte, die erfolgen, wenn der Zuschauer in eine vordefinierte Region schaut
spatial	räumlich, z.B. räumlicher Ton in einer VR-Welt
spatio-temporale Daten	Daten mit Raum- und Zeitdimensionen (Raumdaten, die sich zeitlich ändern)
sphärische Filme	Filme, die auf das innere einer Kugel projiziert werden. Mittels VR-Geräten ist es dem Zuschauer möglich, den Film aus dem Inneren dieser Kugel zu sehen und dadurch das Gefühl zu haben, vor Ort zu sein.
SGD - Subtil Gaze Direction	Lenkung des Blickes, ohne dass sich der Zuschauer dieser Lenkung bewusst ist, spezielle Methode von Bailey mittels Flicker-Effekten (Bailey u. a., 2009)
subtil	unbemerkt, nicht oder kaum wahrnehmbar
Szene	Einheit eines Films mit Kontinuität in Zeit, Ort und Charakteren
Target	zu aktivierendes Ziel
Tiefencues	Hinweise zur Wahrnehmung räumlicher Tiefe (depth cues), dienen der Einschätzung räumlicher Entfernungen
Timeline	Komponente traditioneller Schnittprogrammen, auf der die einzelnen Filmsegmente in ihrer zeitlichen Abfolge und Länge angeordnet werden
timelinebasierte Schnitte	Schnitte, die zu einem vordefinierten Zeitpunkt (Timecode) erfolgen
Track	Koordinaten zurückgelegter Wege. Dies können die Blick- oder Kopfrichtungen sein, aber auch Positionen von Teilnehmern oder Controllern in einer VR-Umgebung.
Transition	Technik mittels der von einer Einstellung zur nächsten gewechselt wird
Trigger	Auslöser für ein Ereignis
Usability	Benutzbarkeit, Benutzerfreundlichkeit, Gebrauchstauglichkeit
User Experience	Nutzererlebnis, Eindrücke des Nutzers bei der Interaktion
Viewport	Ausschnitt eines Bildes, welcher auf einem Display sichtbar ist
VR – Virtual Reality	virtuelle Realität, die es ermöglicht, den Benutzer an einen virtuellen Ort zu transportieren und ihm das Gefühl gibt, dort tatsächlich präsent zu sein.

VR-Gerät	Gerät, mit dem der Nutzer in die virtuelle Realität eintauchen kann, z.B. HMD, CAVE, Cardboards
world-referenziert	mit der virtuellen Welt verbunden, in CVR mit dem sphärischen Film. World-referenzierte Objekte bleiben an ihrem Platz in der Filmwelt fixiert, auch wenn der Betrachter den Kopf bewegt.
yaw / Gieren	Drehung um die z-Achse eines Referenzsystems

Literaturverzeichnis

- Abreu, J., Almeida, P., Branco, V., 2002. 2BeOn - Interactive Television Supporting Interpersonal Communication, in: *Multimedia 2001. Eurographics*. Springer, Vienna, S. 199–208.
https://doi.org/10.1007/978-3-7091-6103-6_20
- Agulló, B., Montagud, M., Fraile, I., 2019. Making Interaction with Virtual Reality Accessible: Rendering and Guiding Methods for Subtitles. *Intelligent Interaction Design - AI EDAM* 33(4), 416–428.
<https://doi.org/10.1017/S0890060419000362>
- Aitamurto, T., Zhou, S., Sakshuwong, S., Saldivar, J., Sadeghi, Y., Tran, A., 2018. Sense of Presence, Attitude Change, Perspective-Taking and Usability in First-Person Split-Sphere 360° Video, in: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM, New York, NY, USA, S. 1–12. <https://doi.org/10.1145/3173574.3174119>
- Akahori, W., Hirai, T., Kawamura, S., Morishima, S., 2016. Region-of-Interest-Based Subtitle Placement Using Eye-Tracking Data of Multiple Viewers, in: *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '16*. ACM, New York, NY, USA, S. 123–128. <https://doi.org/10.1145/2932206.2933558>
- Altenloh, E., 1913. *Zur Soziologie des Kino*. PhD Thesis, Ruprecht-Karls-Universität zu Heidelberg.
- Amerson, D., Kime, S., Young, R.M., 2005. Real-time Cinematic Camera Control for Interactive Narratives, in: *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology - ACE '05*. ACM, New York, NY, USA, S. 369–369.
<https://doi.org/10.1145/1178477.1178552>
- Argelaguet, F., Andujar, C., 2013. A Survey of 3D Object Selection Techniques for Virtual Environments. *Computers & Graphics* 37(3), 121–136. <https://doi.org/10.1016/J.CAG.2012.12.003>
- Argyriou, L., Economou, D., Bouki, V., Doumanis, I., 2016. Engaging Immersive Video Consumers: Challenges Regarding 360-Degree Gamified Video Applications, in: *2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security - IUCC-CSS*. IEEE, S. 145–152. <https://doi.org/10.1109/IUCC-CSS.2016.028>
- Arijon, D., 2000. *Grammatik der Filmsprache: Das Handbuch*. Zweitausendeins, Frankfurt am Main.
- Arijon, D., 1991. *Grammar of the Film Language*. Silman-James Press.
- Asher, T., Ogle, E., Bailenson, J., Herrera, F.F., 2018. Becoming homeless. *Association for Computing Machinery (ACM)*, S. 1–1. <https://doi.org/10.1145/3226552.3226576>
- Atienza, R., Blonna, R., Saldares, M.I., Casimiro, J., Fuentes, V., 2016. Interaction Techniques Using Head Gaze for Virtual Reality, in: *2016 IEEE Region 10 Symposium (TENSYP)*. IEEE, S. 110–114.
<https://doi.org/10.1109/TENCONSpring.2016.7519387>
- Bach, B., Dragicevic, P., Archambault, D., Hurter, C., Carpendale, S., 2014. A Review of Temporal Data Visualizations Based on Space-Time Cube Operations, in: *Eurographics Conference on Visualization - EUROVIS*, hal-01006140. Swansea, Wales, United Kingdom, S. 1–19.
- Bailenson, J.N., Blascovich, J., Beall, A.C., Loomis, J.M., 2001. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 10(6), 583–598. <https://doi.org/10.1162/105474601753272844>
- Bailey, R., McNamara, A., Costello, A., Sridharan, S., Grimm, C., 2012. Impact of Subtle Gaze Direction on Short-Term Spatial Information Recall - ETRA '12, in: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, New York, NY, USA, S. 67–74.
<https://doi.org/10.1145/2168556.2168567>
- Bailey, R., McNamara, A., Sudarsanam, N., Grimm, C., 2009. Subtle Gaze Direction. *ACM Transactions on*

- Graphics 28(4), 1–14. <https://doi.org/10.1145/1559755.1559757>
- Balázs, B., 1924. *Der sichtbare Mensch oder die Kultur des Films*. Deutsch-Österreichischer Verlag, Wien-Leipzig.
- Ball, K.K., Beard, B.L., Roenker, D.L., Miller, R.L., Griggs, D.S., 1988. Age and Visual Search: Expanding the Useful Field of View. *Journal of the Optical Society of America A* 5(12), 2210. <https://doi.org/10.1364/JOSAA.5.002210>
- Barth, E., Dorr, M., Böhme, M., Gegenfurtner, K., Martinetz, T., 2006. Guiding the Mind's Eye: Improving Communication and Vision by External Control of the Scanpath, in: *Proceedings SPIE 6057, Human Vision and Electronic Imaging XI*. SPIE, S. 116–123. <https://doi.org/10.1117/12.674147>
- Baudisch, P., Rosenholtz, R., 1993. Halo: A Technique for Visualizing Off-Screen Locations, in: *Proceedings of the Conference on Human Factors in Computing Systems - CHI '03*. ACM, New York, NY, USA, S. 481–488. <https://doi.org/10.1145/642611.642695>
- Belda, J., Montagud M., Boronat, F., Martinez, M., Pastor, J., 2015. Wersync: A WEB-Based Platform for Distributed Media Synchronization and Social Interaction, in: *Proceedings of ACM International Conference on Interactive Experiences for Television and Online Video 2015*. ACM, New York, NY, USA, S. 9–10.
- Beller, H., 2002. *Handbuch der Filmmontage, Praxis und Prinzipien des Filmschnitts*. TR-Verlagsunion, München.
- Biocca, F., Owen, C., Tang, A., Bohil, C., 2007. Attention Issues in Spatial Information Systems: Directing Mobile Users' Visual Attention Using Augmented Reality. *Journal of Management Information Systems* 23(4), 163–184. <https://doi.org/10.2753/MIS0742-1222230408>
- Biocca, F., Tang, A., Owen, C., Xiao, F., 2006. Attention Funnel: Omnidirectional 3D Cursor for Mobile Augmented Reality Platforms, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '06*. ACM, New York, NY, USA, S. 1115–1122. <https://doi.org/10.1145/1124772.1124939>
- Blascheck, T., Kurzahls, K., Raschke, M., Burch, M., Weiskopf, D., Ertl, T., 2017. Visualization of Eye Tracking Data: A Taxonomy and Survey. *Computer Graphics Forum* 36(8), 260–284. <https://doi.org/10.1111/cgf.13079>
- Bloom, P., 2017a. Empathy and Its Discontents. *Trends in Cognitive Sciences* 21(1), 24–31. <https://doi.org/10.1016/J.TICS.2016.11.004>
- Bloom, P., 2016. *Against Empathy: The Case for Rational Compassion*. Ecco.
- Boonsuk, W., 2011. *Evaluation of Desktop Interface Displays for 360-Degree Video*. Thesis, Iowa State University, Digital Repository, Ames. <https://doi.org/10.31274/etd-180810-2797>
- Booth, T., Sridharan, S., McNamara, A., Grimm, C., Bailey, R., 2013. Guiding Attention in Controlled Real-World Environments, in: *Proceedings of the ACM Symposium on Applied Perception - SAP '13*. ACM, New York, NY, USA, S. 75–82. <https://doi.org/10.1145/2492494.2492508>
- Bordwell, D., Thompson, K., 2013. *Film Art : An Introduction*. McGraw-Hill.
- Borji, A., Itti, L., 2013. State-of-the-Art in Visual Attention Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(1), 185–207. <https://doi.org/10.1109/TPAMI.2012.89>
- Bork, F., Schnelzer, C., Eck, U., Navab, N., 2018. Towards Efficient Visual Guidance in Limited Field-of-View Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 24(11), 2983–2992. <https://doi.org/10.1109/TVCG.2018.2868584>
- Boronat, F., Montagud, M., Marfil, D., Luzon, C., 2018. Hybrid Broadcast/Broadband TV Services and Media Synchronization: Demands, Preferences and Expectations of Spanish Consumers. *IEEE Transactions on Broadcasting* 64(1), 52–69. <https://doi.org/10.1109/TBC.2017.2737819>

- Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I., 2005. 3D User Interfaces: Theory and Practice. Addison-Wesley.
- Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I., 2001. An Introduction to 3-D User Interface Design. Presence: Teleoperators and Virtual Environments 10(1), 96–108.
<https://doi.org/10.1162/105474601750182342>
- Brooks, M., Armstrong, M., 2014. Enhancing Subtitles, in: Short Papers of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '14. ACM, New York, NY, USA, S. 25–27.
- Brown, A., Jones, R., Crabb, M., Sandford, J., Brooks, M., Armstrong, M., Jay, C., 2015. Dynamic Subtitles: The User Experience, in: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '15. ACM, New York, NY, USA, S. 103–112.
<https://doi.org/10.1145/2745197.2745204>
- Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., Glancy, M., 2017. Subtitles in 360-degree Video, in: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '17. ACM, New York, NY, USA, S. 3–8.
<https://doi.org/10.1145/3084289.3089915>
- Brown, C., Bhutra, G., Suhail, M., Xu, Q., Ragan, E.D., 2017. Coordinating Attention and Cooperation in Multi-user Virtual Reality Narratives, in: 2017 IEEE Virtual Reality Conference - IEEEVR '17. IEEE, S. 377–378. <https://doi.org/10.1109/VR.2017.7892334>
- Buccalo, W.R., 1977. Mise-en-Scene Versus Montage: Viewer Response to Two Styles of Visual Communication. Ph.D. Thesis, The Ohio State University.
- Burch, M., Kull, A., Weiskopf, D., 2013. AOI Rivers for Visualizing Dynamic Eye Gaze Frequencies. Computer Graphics Forum 32(3pt3), 281–290. <https://doi.org/10.1111/cgf.12115>
- Burgos-Artizzu, X.P., Fleureau, J., Dumas, O., Tapie, T., LeClerc, F., Mollet, N., 2015. Real-time Expression-sensitive HMD Face Reconstruction, in: SIGGRAPH ASIA 2015 Technical Briefs on - SA '15. ACM, New York, NY, USA, S. 1–4. <https://doi.org/10.1145/2820903.2820910>
- Burigat, S., Chittaro, L., Gabrielli, S., 2006. Visualizing Locations of Off-Screen Objects on Mobile Devices: A Comparative Evaluation of Three Approaches, in: Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '06. ACM, New York, NY, USA, S. 239–246. <https://doi.org/10.1145/1152215.1152266>
- Carlsson, C., Hagsand, O., 1993. DIVE - a Multi-User Virtual Reality System, in: Proceedings of IEEE Virtual Reality Annual International Symposium, 1993. IEEE, S. 394–400.
<https://doi.org/10.1109/VRAIS.1993.380753>
- Carrasco, M., 2011. Visual Attention: The Past 25 Years. Vision Research 51(13), 1484–1525.
<https://doi.org/10.1016/J.VISRES.2011.04.012>
- Chambel, T., Chhaganlal, M.N., Neng, L.A.R., 2011. Towards Immersive Interactive Video. Through 360° Hypervideo, in: Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology - ACE '11. ACM, New York, NY, USA, S. 1–2.
<https://doi.org/10.1145/2071423.2071518>
- Chang, H.-Y., Tseng, W.-J., Tsai, C.-E., Chen, H.-Y., Peiris, R.L., Chan, L., 2018. FacePush: Introducing Normal Force on Face with Head-Mounted Displays, in: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18. ACM, New York, NY, USA, S. 927–935. <https://doi.org/10.1145/3242587.3242588>
- Chauvin, S., Levieux, G., Donnart, J.-Y., Natkin, S., 2014. An Out-of-Character Approach to Emergent Game Narratives, in: Proceedings of the 9th International Conference on the Foundations of Digital Games, hal-01126515. S. 1–4.

- Chen, H., Yan, M., Liu, S., Jiang, B., 2017. Gaze Inspired Subtitle Position Evaluation for MOOCs Videos, in: Proceedings SPIE 10443, Second International Workshop on Pattern Recognition. SPIE, S. 216–220. <https://doi.org/10.1117/12.2280281>
- Choe, M., Choi, Y., Park, J., Kim, H.K., 2018. Comparison of Gaze Cursor Input Methods for Virtual Reality Devices. *International Journal of Human–Computer Interaction* 35(7), 620–629. <https://doi.org/10.1080/10447318.2018.1484054>
- Choumane, A., Casiez, G., Grisoni, L., 2010. Buttonless Clicking: Intuitive Select and Pick-release Through Gesture Analysis, in: 2010 IEEE Virtual Reality Conference - IEEEVR '10. IEEE, S. 67–70. <https://doi.org/10.1109/VR.2010.5444810>
- Clarke, K.C., 2017. Map Projections and Coordinate Systems, in: Richardson, D., Castree, N., Goodchild, M.F., Kobayashi, A., Liu, W., Marston, R.A. (Hrsg.), *International Encyclopedia of Geography: People, the Earth, Environment and Technology*. John Wiley & Sons, Ltd, Oxford, UK, S. 1–10. <https://doi.org/10.1002/9781118786352.wbieg0455>
- Cole, F., DeCarlo, D., Finkelstein, A., Kin, K., Morley, K., Santella, A., 2006. Directing Gaze in 3D Models with Stylized Focus, in: Proceedings of the 17th Eurographics Conference on Rendering Techniques - EGSR '06. Eurographics Association, S. 377–387. <https://doi.org/10.2312/egwr/egsr06/377-387>
- Corbillon, X., De Simone, F., Simon, G., 2017. 360-Degree Video Head Movement Dataset, in: Proceedings of the 8th ACM on Multimedia Systems Conference - MMSys '17. ACM, New York, NY, USA, S. 199–204. <https://doi.org/10.1145/3193701>
- Cordeil, M., Dwyer, T., Klein, K., Laha, B., Marriott, K., Thomas, B.H., 2017. Immersive Collaborative Analysis of Network Connectivity: CAVE-style or Head-Mounted Display? *IEEE Transactions on Visualization and Computer Graphics* 23(1), 441–450. <https://doi.org/10.1109/TVCG.2016.2599107>
- Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R. V., Hart, J.C., 1992. The CAVE: Audio Visual Experience Automatic Virtual Environment. *Communications of the ACM* 35(6), 64–72. <https://doi.org/10.1145/129888.129892>
- Cutting, J.E., 2014. Event Segmentation and Seven Types of Narrative Discontinuity in Popular Movies. *Acta Psychologica* 149, 69–77. <https://doi.org/10.1016/J.ACTPSY.2014.03.003>
- Cutting, J.E., 2007. Perceiving Scenes in Film and in the World, in: Anderson, J.D., Fisher Anderson, B. (Hrsg.), *Moving image theory: Ecological considerations*. Southern Illinois University Press, S. 9–27. <https://doi.org/use.jhu.edu/book/24943>
- d'Ydewalle, G., Desmet, G., Van Rensbergen, J., 1998. Film Perception: The Processing of Film Cuts. *Eye Guidance in Reading and Scene Perception* 357–367. <https://doi.org/10.1016/B978-008043361-5/50017-1>
- Danieau, F., Guillo, A., Dore, R., 2017. Attention Guidance for Immersive Video Content in Head-Mounted Displays, in: 2017 IEEE Virtual Reality - IEEEVR '17. IEEE, S. 205–206. <https://doi.org/10.1109/VR.2017.7892248>
- Davies, A.G., Crohn, N.J., Treadgold, L.A., 2019. Can Virtual Reality Really be Used within the Lecture Theatre? *BMJ Simulation and Technology Enhanced Learning* 5(4), 234–235. <https://doi.org/10.1136/bmjstel-2017-000295>
- Davis, S., Nesbitt, K., Nalivaiko, E., 2014. A Systematic Review of Cybersickness, in: Proceedings of the 2014 Conference on Interactive Entertainment - IE2014. ACM, New York, NY, USA, S. 1–9. <https://doi.org/10.1145/2677758.2677780>
- De Abreu, A., Ozcinar, C., Smolic, A., 2017. Look Around You: Saliency Maps for Omnidirectional Images in VR Applications, in: 2017 Ninth International Conference on Quality of Multimedia Experience - QoMEX. IEEE, S. 1–6. <https://doi.org/10.1109/QoMEX.2017.7965634>

- De Greef, P., IJsselstein, W.A., 2001. Social Presence in a Home Tele-Application. *CyberPsychology & Behavior* 4(2), 307–315. <https://doi.org/10.1089/109493101300117974>
- Dorr, M., Vig, E., Barth, E., 2012. Eye Movement Prediction and Variability on Natural Video Data Sets. *Visual Cognition* 20(4–5), 495–514. <https://doi.org/10.1080/13506285.2012.667456>
- Dorta, T., Pierini, D., Boudhraâ, S., 2016. Why 360° and VR Headsets for Movies?: Exploratory Study of Social VR via Hyve-3D, in: *Actes de la 28ième conférence francophone sur l'Interaction Homme-Machine*. ACM, New York, NY, USA, S. 211–220. <https://doi.org/10.1145/3004107.3004117>
- Dreyfuss, H., 1967. *The Measure of Man: Human Factors in Design*, The American journal of sports medicine. Whitney Library of Design New York, New York.
- Drusch, G., Bastien, J.M.C., Paris, S., 2014. Analysing Eye-tracking Data: from Scanpaths and Heatmaps to the Dynamic Visualisation of Areas of Interest, in: *International Conference on Applied Human Factors and Ergonomics*, hal-01223743. Krakow, Poland, S. 205–211.
- Durlach, N., Slater, M., 2000. Presence in Shared Virtual Environments and Virtual Togetherness. *Presence: Teleoperators and Virtual Environments* 9(2), 214–217. <https://doi.org/10.1162/105474600566736>
- Evans, C., Gibbons, N.J., 2007. The Interactivity Effect in Multimedia Learning. *Computers and Education* 49(4), 1147–1160. <https://doi.org/10.1016/j.compedu.2006.01.008>
- Fearghail, C.O., Ozcinar, C., Knorr, S., Smolic, A., 2018. Director's Cut - Analysis of VR Film Cuts for Interactive Storytelling, in: *2018 International Conference on 3D Immersion - IC3D*. IEEE, S. 1–8. <https://doi.org/10.1109/IC3D.2018.8657901>
- Fetter, M., 2007. Vestibulo-Ocular Reflex, in: Straube, A., Büttner, U. (Hrsg.), *Neuro-Ophthalmology*. Karger, Basel, S. 35–51. <https://doi.org/10.1159/000100348>
- Fidyka, A., Matamala, A., 2018. Audio description in 360° videos. *Translation Spaces* 7(2), 285–303. <https://doi.org/10.1075/ts.18018.fid>
- Fisher, J.A., 2017. Empathic Actualities: Toward a Taxonomy of Empathy in Virtual Reality, in: *International Conference on Interactive Digital Storytelling - ICIDS*. Springer, Cham, S. 233–244. https://doi.org/10.1007/978-3-319-71027-3_19
- Freeman, D., Evans, N., Lister, R., Antley, A., Dunn, G., Slater, M., 2014. Height, Social Comparison, and Paranoia: An Immersive Virtual Reality Experimental Study. *Psychiatry Research* 218(3), 348–352. <https://doi.org/10.1016/J.PSYCHRES.2013.12.014>
- Frintrop, S., Rome, E., Christensen, H.I., 2010. Computational Visual Attention Systems and Their Cognitive Foundations: A Survey. *ACM Transactions on Applied Perception* 7(1), 1–39. <https://doi.org/10.1145/1658349.1658355>
- Galvane, Q., Ronfard, R., Christie, M., Szilas, N., 2014. Narrative-Driven Camera Control for Cinematic Replay of Computer Games, in: *Proceedings of the Seventh International Conference on Motion in Games - MIG '14*. ACM Press, New York, New York, USA, S. 109–117. <https://doi.org/10.1145/2668064.2668104>
- Geerts, D., Vaishnavi, I., Mekuria, R., Van Deventer, O., Cesar, P., 2011. Are we in Sync? Synchronization Requirements for Watching Online Video Together, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '11*. ACM, New York, NY, USA, S. 311–314. <https://doi.org/10.1145/1978942.1978986>
- Getis, A., Ord, J.K., 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* 24(3), 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- Godon, M., Handekyn, K., Vanparijs, F., Coppens, T., 2004. Enabling rich communication with community support for television. *Journal-Communications Network* 3(3), 11–15.

- Grimes, J.D., 1983. Effects of Patterning on Flicker Frequency, in: Proceedings of the Human Factors Society Annual Meeting. SAGE Publications, Los Angeles, CA, S. 46–50.
<https://doi.org/10.1177/154193128302700112>
- Grogorick, S., Stengel, M., Eisemann, E., Magnor, M., 2017. Subtle Gaze Guidance for Immersive Environments, in: Proceedings of the ACM Symposium on Applied Perception - SAP '17. ACM, New York, NY, USA, S. 1–7. <https://doi.org/10.1145/3119881.3119890>
- Gruenefeld, U., El Ali, A., Boll, S., Heuten, W., 2018a. Beyond Halo and Wedge, in: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '18. ACM, New York, NY, USA, S. 1–11. <https://doi.org/10.1145/3229434.3229438>
- Gruenefeld, U., El Ali, A., Heuten, W., Boll, S., 2017a. Visualizing Out-of-view Objects in Head-mounted Augmented Reality, in: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17. ACM, New York, NY, USA, S. 1–7. <https://doi.org/10.1145/3098279.3122124>
- Gruenefeld, U., Ennenga, D., El Ali, A., Heuten, W., Boll, S., 2017b. EyeSee360, in: Proceedings of the 5th Symposium on Spatial User Interaction - SUI '17. ACM, New York, NY, USA, S. 109–118. <https://doi.org/10.1145/3131277.3132175>
- Gruenefeld, U., Stratmann, T.C., El Ali, A., Boll, S., Heuten, W., 2018b. RadialLight, in: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '18. ACM, New York, NY, USA, S. 1–6. <https://doi.org/10.1145/3229434.3229437>
- Gugenheimer, J., Dobbelsstein, D., Winkler, C., Haas, G., Rukzio, E., 2016a. FaceTouch: Enabling Touch Interaction in Display Fixed UIs for Mobile Virtual Reality, in: Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST '16. ACM, New York, NY, USA, S. 49–60. <https://doi.org/10.1145/2984511.2984576>
- Gugenheimer, J., Stemasov, E., Frommel, J., Rukzio, E., 2017. ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17. ACM, New York, NY, USA, S. 4021–4033. <https://doi.org/10.1145/3025453.3025683>
- Gugenheimer, J., Wolf, D., Haas, G., Krebs, S., Rukzio, E., 2016b. SwiVRChair: A Motorized Swivel Chair to Nudge Users' Orientation for 360 Degree Storytelling in Virtual Reality, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16. ACM, New York, NY, USA, S. 1996–2000. <https://doi.org/10.1145/2858036.2858040>
- Guilford, J.P., Helson, H., 1929. Eye-Movements and the Phi-Phenomenon. The American Journal of Psychology 41(4), 595–606. <https://doi.org/10.2307/1414716>
- Gunkel, S.N.B., Prins, M., Stokking, H., Niamut, O., 2017. Social VR Platform: Building 360-degree Shared VR Spaces, in: Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '17 Adjunct. ACM, New York, NY, USA, S. 83–84. <https://doi.org/10.1145/3084289.3089914>
- Gunkel, S.N.B., Stokking, H.M., Prins, M.J., van der Stap, N., ter Haar, F.B., Niamut, O.A., 2018. Virtual Reality Conferencing: Multi-user Immersive VR Experiences on the Web, in: Proceedings of the 9th ACM Multimedia Systems Conference on - MMSys '18. ACM, New York, NY, USA, S. 498–501. <https://doi.org/10.1145/3204949.3208115>
- Gustafson, S., Baudisch, P., Gutwin, C., Irani, P., 2008. Wedge: Clutter-Free Visualization of Off-Screen Locations, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '08. ACM, S. 787–796. <https://doi.org/10.1145/1357054.1357179>
- Gustafson, S.G., Irani, P.P., 2007. Comparing Visualizations for Tracking Off-Screen Moving Targets, in: Extended Abstracts on Human Factors in Computing Systems - CHI '07. ACM, S. 2399–2404. <https://doi.org/10.1145/1240866.1241014>

- Hagiwara, A., Sugimoto, A., Kawamoto, K., 2011. Saliency-Based Image Editing for Guiding Visual Attention, in: Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-Based Interaction - PETMEI '11. ACM, New York, NY, USA, S. 43–48.
<https://doi.org/10.1145/2029956.2029968>
- Hall, E.T., 1966. The Hidden Dimension. Doubleday, New York.
- Hand, C., 1997. A Survey of 3D Interaction Techniques. Computer Graphics Forum 16(5), 269–281.
<https://doi.org/10.1111/1467-8659.00194>
- Hansen, J.P., Rajanna, V., MacKenzie, I.S., Bækgaard, P., 2018. A Fitts' Law Study of Click and Dwell Interaction by Gaze, Head and Mouse with a Head-Mounted Display, in: Proceedings of the Workshop on Communication by Gaze Interaction - COGAIN '18. ACM, New York, NY, USA, S. 1–5.
<https://doi.org/10.1145/3206343.3206344>
- Harboe, G., Massey, N., Metcalf, C., Wheatley, D., Romano, G., 2008a. The Uses of Social Television. Computers in Entertainment 6(1), 1–15. <https://doi.org/10.1145/1350843.1350851>
- Harboe, G., Metcalf, C.J., Bentley, F., Tullio, J., Massey, N., Romano, G., 2008b. Ambient Social TV: Drawing People into a Shared Experience, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '08. ACM, New York, NY, USA, S. 1–10.
<https://doi.org/10.1145/1357054.1357056>
- Harrington, C.M., Kavanagh, D.O., Wright Ballester, G., Wright Ballester, A., Dicker, P., Traynor, O., Hill, A., Tierney, S., 2018. 360° Operative Videos: A Randomised Cross-Over Study Evaluating Attentiveness and Information Retention. Journal of Surgical Education 75(4), 993–1000.
<https://doi.org/10.1016/j.jsurg.2017.10.010>
- Hart, S.G., 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. SAGE Publications, Los Angeles, CA, S. 904–908.
<https://doi.org/10.1177/154193120605000909>
- Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology 52, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hassib, M., Pfeiffer, M., Schneegass, S., Rohs, M., Alt, F., 2017. Emotion Actuator: Embodied Emotional Feedback through Electroencephalography and Electrical Muscle Stimulation, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17. ACM, New York, NY, USA, S. 6133–6146. <https://doi.org/10.1145/3025453.3025953>
- Hata, H., Koike, H., Sato, Y., 2016. Visual Guidance with Unnoticed Blur Effect, in: Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16. ACM, New York, NY, USA, S. 28–35. <https://doi.org/10.1145/2909132.2909254>
- Healey, C.G., Booth, K.S., Enns, J.T., 1996. High-Speed Visual Estimation Using Preattentive Processing. ACM Transactions on Computer-Human Interaction 3(2), 107–135.
<https://doi.org/10.1145/230562.230563>
- Herrera, F., Bailenson, J., Weisz, E., Ogle, E., Zaki, J., 2018. Building Long-term Empathy: A Large-scale Comparison of Traditional and Virtual Reality Perspective-taking. PLOS ONE 13(10), e0204494.
<https://doi.org/10.1371/journal.pone.0204494>
- Ho, C., Basdogan, C., Slater, M., Durlach, N.I., Srinivasan, M.A., 1998. An Experiment on the Influence of Haptic Communication on the Sense of Being Together, in: Proceedings of the British Telecom Workshop on Presence in Shared Virtual Environments. S. 10–11.
- Hoeg, E.R., Gerry, L.J., Thomsen, L., Nilsson, N.C., Serafin, S., 2017. Binaural Sound Reduces Reaction Time in a Virtual Reality Search Task, in: 2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments - SIVE. IEEE, Los Angeles, CA, USA, S. 1–4.
<https://doi.org/10.1109/SIVE.2017.7901610>

- Hong, R., Wang, M., Yuan, X.-T., Xu, M., Jiang, J., Yan, S., Chua, T.-S., 2011. Video Accessibility Enhancement for Hearing-Impaired Users. *ACM Transactions on Multimedia Computing, Communications, and Applications* 7S(1), 1–19. <https://doi.org/10.1145/2037676.2037681>
- Hong, S., Kim, G.J., 2016. Accelerated Viewpoint Panning with Rotational Gain in 360 Degree Videos, in: *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology - VRST '16*. ACM, New York, NY, USA, S. 303–304. <https://doi.org/10.1145/2993369.2996309>
- Hong, Y., MacQuarrie, A., Steed, A., 2018. The Effect of Chair Type on Users' Viewing Experience for 360-degree Video, in: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology - VRST '18*. ACM, New York, NY, USA, S. 1–11. <https://doi.org/10.1145/3281505.3281519>
- Hoppe, M., Knierim, P., Kosch, T., Funk, M., Futami, L., Schneegass, S., Henze, N., Schmidt, A., Machulla, T., 2018. VRHapticDrones: Providing Haptics in Virtual Reality through Quadcopters, in: *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia - MUM '18*. ACM, New York, NY, USA, S. 7–18. <https://doi.org/10.1145/3282894.3282898>
- Hossain, Z., Hasan, K., Liang, H.-N., Irani, P., 2012. EdgeSplit: Facilitating the Selection of Off-Screen Objects, in: *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '12*. ACM, New York, NY, USA, S. 79–82. <https://doi.org/10.1145/2371574.2371588>
- Howard, S., Serpanchy, K., Lewin, K., 2018. Virtual Reality Content for Higher Education Curriculum, in: *Proceedings of the 19th Biennial Conference and Exhibition. Technology and the Future Inc.*, Melbourne, Australia, S. 1–15.
- Hu, Y., Kautz, J., Yu, Y., Wang, W., 2015. Speaker-Following Video Subtitles. *ACM Transactions on Multimedia Computing, Communications, and Applications* 11(2), 1–17. <https://doi.org/10.1145/2632111>
- Hurbis-Cherrier, M., 2018. *Voice and Vision: A Creative Approach to Narrative Filmmaking*, 3. ed. Routledge, New York. <https://doi.org/10.4324/9781315815893>
- IJsselstein, W., van Baren, J., Markopoulos, P., Romero, N., de Ruyter, B., 2009. Measuring Affective Benefits and Costs of Mediated Awareness: Development and Validation of the ABC-Questionnaire, in: *Awareness Systems*. Springer, London, S. 473–488. https://doi.org/10.1007/978-1-84882-477-5_20
- IJsselstein, W.A., de Ridder, H., Freeman, J., Avons, S.E., 2000. Presence: Concept, Determinants and Measurement, in: Rogowitz, B.E., Pappas, T.N. (Hrsg.), *International Society for Optics and Photonics*, S. 520–529. <https://doi.org/10.1117/12.387188>
- Itti, L., Koch, C., 2001. Computational Modelling of Visual Attention. *Nature Reviews Neuroscience* 2(3), 194–203. <https://doi.org/10.1038/35058500>
- Itti, L., Koch, C., Niebur, E., 1998. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(11), 1254–1259. <https://doi.org/10.1109/34.730558>
- Iwerks, D., 2019. *Walt Disney's Ultimate Inventor: The Genius of Ub Iwerks*. Disney Press.
- Jacob, R.J.K., 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At is What You Get. *ACM Transactions on Information Systems* 9(2), 152–169. <https://doi.org/10.1145/123078.128728>
- Jacob, R.J.K., 1990. What You Look At is What You Get: Eye Movement-Based Interaction Techniques, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Empowering People - CHI '90*. ACM, New York, NY, USA, S. 11–18. <https://doi.org/10.1145/97243.97246>
- Jaekl, P.M., Jenkin, M.R., Harris, L.R., 2005. Perceiving a Stable World During Active Rotational and

- Translational Head Movements. *Experimental Brain Research* 163(3), 388–399.
<https://doi.org/10.1007/s00221-004-2191-8>
- Jay, C., Hubbard, R., 2003. Amplifying Head Movements with Head-Mounted Displays. *Presence: Teleoperators and Virtual Environments* 12(3), 268–276.
<https://doi.org/10.1162/105474603765879521>
- Jerald, J., 2015. *The VR Book: Human-Centered Design for Virtual Reality*. Association for Computing Machinery and Morgan & Claypool. <https://doi.org/10.1145/2792790>
- Jhala, A., Young, R.M., 2005. A Discourse Planning Approach to Cinematic Camera Control for Narratives in Virtual Environments, in: *Proceedings of the 20th National Conference on Artificial Intelligence - Volume 1*. AAAI, S. 307–312.
- Jiang, W., Li, J., 2014. The Effects of Spatial Reference Systems on the Predictive Accuracy of Spatial Interpolation Methods. *Geoscience Australia*, Canberra.
- Jo, H., Hwang, S., Park, H., Ryu, J., 2011. Aroundplot: Focus+Context Interface for Off-Screen Objects in 3D Environments. *Computers & Graphics* 35(4), 841–853.
<https://doi.org/10.1016/J.CAG.2011.04.005>
- Kachkovski, G. V., Vasilyev, D., Kuk, M., Kingstone, A., Street, C.N.H., 2019. Exploring the Effects of Violating the 180-Degree Rule on Film Viewing Preferences. *Communication Research* 46(7), 948–964. <https://doi.org/10.1177/0093650219838959>
- Kallioniemi, P., Keskinen, T., Mäkelä, V., Karhu, J., Ronkainen, K., Nevalainen, A., Hakulinen, J., Turunen, M., 2018. Hotspot Interaction in Omnidirectional Videos Using Head-Mounted Displays, in: *Proceedings of the 22nd International Academic Mindtrek Conference - Mindtrek '18*. ACM, New York, NY, USA, S. 126–134. <https://doi.org/10.1145/3275116.3275148>
- Kallioniemi, P., Mäkelä, V., Saarinen, S., Turunen, M., Winter, Y., Istudor, A., 2017. User Experience and Immersion of Interactive Omnidirectional Videos in CAVE Systems and Head-Mounted Displays, in: *IFIP Conference on Human-Computer Interaction - INTERACT 2017*. Springer, Cham, S. 299–318.
https://doi.org/10.1007/978-3-319-68059-0_20
- Kandorfer, P., 2003. *Lehrbuch der Filmgestaltung: Theoretisch-technische Grundlagen der Filmkunde*. Mediabook-Verlag.
- Katz, S.D., 2000. *Die richtige Einstellung: Shot by shot- Zur Bildsprache des Films*. Zweitausendeins.
- Kaul, O.B., Rohs, M., 2017. HapticHead: A Spherical Vibrotactile Grid around the Head for 3D Guidance in Virtual and Augmented Reality, in: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM, New York, NY, USA, S. 3729–3740.
<https://doi.org/10.1145/3025453.3025684>
- Kendon, A., 2010. Spacing and Orientation in Co-present Interaction, in: Esposito A., Campbell N., Vogel C., Hussain A., Nijholt A. (Hrsg.), *Development of Multimodal Interfaces: Active Listening and Synchrony*. Springer, Berlin, Heidelberg, S. 1–15. https://doi.org/10.1007/978-3-642-12397-9_1
- Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G., 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3
- Kim, J., Song, H., Lee, S., 2018. Extrovert and Lonely Individuals' Social TV Viewing Experiences: A Mediating and Moderating Role of Social Presence. *Mass Communication and Society* 21(1), 50–70. <https://doi.org/10.1080/15205436.2017.1350715>
- King, C.M., et al., 1994. Digital Captioning: Effects of Color-Coding and Placement in Synchronized Text-Audio Presentations., in: *Proceedings of ED-MEDIA 94 World Conference on Educational Multimedia and Hypermedia* . S. 329–334.

- Kjær, T., Lillelund, C.B., Moth-Poulsen, M., Nilsson, N.C., Nordahl, R., Serafin, S., 2017. Can You Cut It? An Exploration of the Effects of Editing in Cinematic Virtual Reality, in: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17. ACM, New York, NY, USA, S. 1–4. <https://doi.org/10.1145/3139131.3139166>
- Knierim, P., Kosch, T., Achberger, A., Funk, M., 2018. Flyables: Exploring 3D Interaction Spaces for Levitating Tangibles, in: Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '18. ACM, New York, NY, USA, S. 329–336. <https://doi.org/10.1145/3173225.3173273>
- Knorr, S., Ozcinar, C., Fearghail, C.O., Smolic, A., 2018. Director's Cut- A Combined Dataset for Visual Attention Analysis in Cinematic VR Content, in: Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production - CVMP '18. ACM, New York, NY, USA, S. 1–10. <https://doi.org/10.1145/3278471.3278472>
- Kolasinski, E.M., 1995. Simulator Sickness in Virtual Environments. United States Army Research Institute for the Behavioral and Social Sciences.
- Koskinen, E., Rakkolainen, I., Raisamo, R., 2017. Direct retinal signals for virtual environments, in: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17. ACM, New York, NY, USA, S. 1–2. <https://doi.org/10.1145/3139131.3141197>
- Kruger, J.-L., Hefer, E., Matthew, G., 2013. Measuring the Impact of Subtitles on Cognitive Load: Eye Tracking and Dynamic Audiovisual Texts, in: Proceedings of the 2013 Conference on Eye Tracking South Africa - ETSA '13. ACM, New York, NY, USA, S. 62–66. <https://doi.org/10.1145/2509315.2509331>
- Kurzahls, K., Cetinkaya, E., Hu, Y., Wang, W., Weiskopf, D., 2017. Close to the Action: Eye-Tracking Evaluation of Speaker-Following Subtitles, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17. ACM, New York, NY, USA, S. 6559–6568. <https://doi.org/10.1145/3025453.3025772>
- Kurzahls, K., Weiskopf, D., 2013. Space-Time Visual Analytics of Eye-Tracking Data for Dynamic Stimuli. IEEE Transactions on Visualization and Computer Graphics 19(12), 2129–2138. <https://doi.org/10.1109/TVCG.2013.194>
- Langbehn, E., Wittig, J., Katzakis, N., Steinicke, F., 2019. Turn Your Head Half Round: VR Rotation Techniques for Situations With Physically Limited Turning Angle, in: Proceedings of Mensch und Computer 2019 - MuC '19. ACM, New York, NY, USA, S. 235–243. <https://doi.org/10.1145/3340764.3340778>
- Laugwitz, B., Held, T., Schrepp, M., 2008. Construction and Evaluation of a User Experience Questionnaire, in: Symposium of the Austrian HCI and Usability Engineering Group. Springer, Berlin, Heidelberg, S. 63–76. https://doi.org/10.1007/978-3-540-89350-9_6
- Lescop, L., Maver, T., Chapman, P., Platt, C., Portela, V., Eaton, D., 2017. 360° Vision, From Panoramas to VR, in: 13th Biennial International Conference of the European Architectural Envisioning Association at the Glasgow School of Art, hal-01584883. Glasgow, S. 226–232.
- Leyrer, M., Linkenauger, S.A., Bühlhoff, H.H., Kloos, U., Mohler, B., 2011. The Influence of Eye Height and Avatars on Egocentric Distance Estimates in Immersive Virtual Environments, in: Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization - APGV '11. ACM, New York, NY, USA, S. 67–74. <https://doi.org/10.1145/2077451.2077464>
- Leyrer, M., Linkenauger, S.A., Bühlhoff, H.H., Mohler, B.J., 2015. The Importance of Postural Cues for Determining Eye Height in Immersive Virtual Reality. PLOS ONE 10(5), e0127000. <https://doi.org/10.1371/journal.pone.0127000>
- Li, F., Bai, H., Zhao, Y., 2017. Visual Attention Guided Eye Movements for 360 Degree Images, in: 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference -

- APSIPA ASC. IEEE, Kuala Lumpur, S. 506–511. <https://doi.org/10.1109/APSIPA.2017.8282085>
- Lin, Y.-C., Chang, Y.-J., Hu, H.-N., Cheng, H.-T., Huang, C.-W., Sun, M., 2017. Tell Me Where to Look: Investigating Ways for Assisting Focus in 360° Video, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17. ACM, New York, NY, USA, S. 2535–2545. <https://doi.org/10.1145/3025453.3025757>
- Lin, Y.-T., Liao, Y.-C., Teng, S.-Y., Chung, Y.-J., Chan, L., Chen, B.-Y., 2017. Outside-In: Visualizing Out-of-Sight Regions-of-Interest in a 360 Video Using Spatial Picture-in-Picture Previews, in: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST '17. ACM, New York, NY, USA, S. 255–265. <https://doi.org/10.1145/3126594.3126656>
- Llobera, J., Spanlang, B., Ruffini, G., Slater, M., 2010. Proxemics with Multiple Dynamic Characters in an Immersive Virtual Environment. *ACM Transactions on Applied Perception* 8(1), 1–12. <https://doi.org/10.1145/1857893.1857896>
- Loomis, J.M., Knapp, J.M., 2003. Visual Perception of Egocentric Distance in Real and Virtual Environments 35–60. <https://doi.org/10.1201/9781410608888-6>
- MacQuarrie, A., Steed, A., 2018. The Effect of Transition Type in Multi-View 360° Media. *IEEE Transactions on Visualization and Computer Graphics* 24(4), 1564–1573. <https://doi.org/10.1109/TVCG.2018.2793561>
- Madary, M., Metzinger, T.K., 2016. Recommendations for Good Scientific Practice and the Consumers of VR-Technology. *Frontiers in Robotics and AI* 3, 1–23. <https://doi.org/10.3389/frobt.2016.00003>
- Maeda, M., Tominaga, T., Hijikata, Y., Sakata, N., 2016. Controlling Virtual Body Size to Reduce Discomfort Caused by Inappropriate Interpersonal Distance, in: Proceedings of the 2016 ACM International Symposium on Wearable Computers - ISWC '16. ACM, New York, NY, USA, S. 192–195. <https://doi.org/10.1145/2971763.2971770>
- Magliano, J.P., Zacks, J.M., 2011. The Impact of Continuity Editing in Narrative Film on Event Segmentation. *Cognitive Science* 35(8), 1489–1517. <https://doi.org/10.1111/j.1551-6709.2011.01202.x>
- Majoranta, P., Aula, A., Rähkä, K.-J., 2004. Effects of Feedback on Eye Typing with a Short Dwell Time, in: Proceedings of the Symposium on Eye Tracking Research & Applications - ETRA '04. ACM, New York, NY, USA, S. 139–146. <https://doi.org/10.1145/968363.968390>
- Mardanbegi, D., Hansen, D.W., Pederson, T., 2012. Eye-Based Head Gestures, in: Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12. ACM, New York, NY, USA, S. 139–146. <https://doi.org/10.1145/2168556.2168578>
- Mardanbegi, D., Hansen, W.D., 2012. Eye-Based Head Gestures for Interaction in the Car, in: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - Automotiveui '12. ACM, S. 39–43.
- Margery, D., Arnaldi, B., Plouzeau, N., 1999. A General Framework for Cooperative Manipulation in Virtual Environments, in: *Virtual Environments '99*. Springer, S. 169–178. https://doi.org/10.1007/978-3-7091-6805-9_17
- Marklund, S., 2017. Perception of Characters in VR: How a Character's Height Affects the Perception of that Character. Thesis, School of Computer Science and Communication.
- Marquardt, N., Hinckley, K., Greenberg, S., 2012. Cross-Device Interaction via Micro-mobility and F-formations, in: Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology - UIST '12. ACM, New York, NY, USA, S. 13–22. <https://doi.org/10.1145/2380116.2380121>
- Marshall, P., Rogers, Y., Pantidi, N., 2011. Using F-formations to Analyse Spatial Patterns of Interaction in Physical Environments, in: Proceedings of the ACM 2011 Conference on Computer Supported

- Cooperative Work - CSCW '11. ACM, New York, NY, USA, S. 445–454.
<https://doi.org/10.1145/1958824.1958893>
- Martín-Gutiérrez, J., Mora, C.E., Añorbe-Díaz, B., González-Marrero, A., 2017. Virtual Technologies Trends in Education. *Eurasia Journal of Mathematics, Science and Technology Education* 13(2), 469–486. <https://doi.org/10.12973/eurasia.2017.00626a>
- Mateer, J., 2017. Directing for Cinematic Virtual Reality: how the traditional film director's craft applies to immersive environments and notions of presence. *Journal of Media Practice* 18(1), 14–25.
<https://doi.org/10.1080/14682753.2017.1305838>
- Mckenzie, S., Rough, J., Spence, A., Patterson, N., 2019. Virtually There: The Potential, Process and Problems of Using 360° Video in the Classroom. *Issues in Informing Science and Information Technology* 16, 211–219. <https://doi.org/10.28945/4318>
- McNamara, A., Bailey, R., Grimm, C., 2008. Improving Search Task Performance Using Subtle Gaze Direction, in: *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization - APGV '08*. ACM, New York, NY, USA, S. 51–56.
<https://doi.org/10.1145/1394281.1394289>
- McNamara, A., Booth, T., Sridharan, S., Caffey, S., Grimm, C., Bailey, R., 2012. Directing Gaze in Narrative Art, in: *Proceedings of the ACM Symposium on Applied Perception - SAP '12*. ACM, New York, NY, USA, S. 63–70. <https://doi.org/10.1145/2338676.2338689>
- Medeiros, D., Cordeiro, E., Mendes, D., Sousa, M., Raposo, A., Ferreira, A., Jorge, J., 2016. Effects of Speed and Transitions on Target-based Travel Techniques, in: *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology - VRST '16*. ACM, New York, NY, USA, S. 327–328. <https://doi.org/10.1145/2993369.2996348>
- Men, L., Bryan-Kinns, N., Hassard, A.S., Ma, Z., 2017. The Impact of Transitions on User Experience in Virtual Reality, in: *2017 IEEE Virtual Reality Conference - IEEEVR '17*. IEEE, S. 285–286.
<https://doi.org/10.1109/VR.2017.7892288>
- Milgram, P., Kishino, F., 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems* E77-D(12), 1321–1329.
- Moghadam, K.R., Ragan, E.D., 2017. Towards Understanding Scene Transition Techniques in Immersive 360 Movies and Cinematic Experiences, in: *2017 IEEE Virtual Reality Conference - IEEEVR '17*. IEEE, S. 375–376. <https://doi.org/10.1109/VR.2017.7892333>
- Mohan, P., Goh, W.B., Fu, C.-W., Yeung, S.-K., 2018. DualGaze: Addressing the Midas Touch Problem in Gaze Mediated VR Interaction, in: *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct - ISMAR-Adjunct*. IEEE, S. 79–84. <https://doi.org/10.1109/ISMAR-Adjunct.2018.00039>
- Montagud, M., Boronat, F., Stokking, H., van Brandenburg, R., 2012. Inter-Destination Multimedia Synchronization: Schemes, Use Cases and Standardization. *Multimedia Systems* 18(6), 459–482.
<https://doi.org/10.1007/s00530-012-0278-9>
- Montagud, M., Cesar, P., Boronat, F., Marfil, D., 2015. Social Media Usage Combined with TV/Video Watching: Opportunities and Associated Challenges. *IEEE Computer Society STCSN E-Letter* 3(2), 1–6.
- Montagud, M., Fraile, I., Meyerson, E., Genís, M., Fernández, S., 2018. ImAc: Enabling Immersive, Accessible and Personalized Media Experiences, in: *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '18*. ACM.
- Montagud, M., Orero, P., Matamala, A., 2020. Culture 4 All: Accessibility-Enabled Cultural Experiences through Immersive VR360 Content. *Personal and Ubiquitous Computing*.
<https://doi.org/10.1007/s00779-019-01357-3>

- Moreno, R., Mayer, R.E., 1999. Cognitive Principles of Multimedia Learning: The Role of Modality and Contiguity. *Journal of Educational Psychology* 91(2), 358–368. <https://doi.org/10.1037/0022-0663.91.2.358>
- Mott, M., Cutrell, E., Gonzalez Franco, M., Holz, C., Ofek, E., Stoakley, R., Ringel Morris, M., 2019. Accessible by Design: An Opportunity for Virtual Reality, in: *Adjunct Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality - ISMAR-Adjunct*. IEEE, S. 451–454. <https://doi.org/10.1109/ISMAR-Adjunct.2019.00122>
- Mulcahy, K.A., Clarke, K.C., 2001. Symbolization of Map Projection Distortion: A Review. *Cartography and Geographic Information Science* 28(3), 167–182. <https://doi.org/10.1559/152304001782153044>
- Murch, W., 1992. In the Blink of an Eye. A Perspective on Film Editing, 2. ed. Silman-James Press.
- Nathan, M., Harrison, C., Yarosh, S., Terveen, L., Stead, L., Amento, B., 2008. CollaboraTV: Making Television Viewing Social Again, in: *Proceeding of the 1st International Conference on Designing Interactive User Experiences for TV and Video - UXTV '08*. ACM, New York, NY, USA, S. 85–94. <https://doi.org/10.1145/1453805.1453824>
- Neng, L.A.R., Chambel, T., 2010. Get Around 360° Hypervideo, in: *Proceedings of the 14th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '10*. ACM, New York, NY, USA, S. 119–122. <https://doi.org/10.1145/1930488.1930512>
- Nguyen, C., DiVerdi, S., Hertzmann, A., Liu, F., 2017. CollaVR: Collaborative In-Headset Review for VR Video, in: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST '17*. ACM Press, New York, New York, USA, S. 267–277. <https://doi.org/10.1145/3126594.3126659>
- Nielsen, L.T., Møller, M.B., Hartmeyer, S.D., Ljung, T.C.M., Nilsson, N.C., Nordahl, R., Serafin, S., 2016. Missing The Point: An Exploration of How to Guide Users' Attention During Cinematic Virtual Reality, in: *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology - VRST '16*. ACM, New York, NY, USA, S. 229–232. <https://doi.org/10.1145/2993369.2993405>
- Nilsson, N.C., Serafin, S., Nordahl, R., 2016. Walking in Place Through Virtual Worlds, in: *International Conference on Human-Computer Interaction*. Springer, Cham, S. 37–48. https://doi.org/10.1007/978-3-319-39516-6_4
- Normand, V., Babski, C., Benford, S., Bullock, A., Carion, S., Chrysanthou, Y., Farcet, N., Frécon, E., Harvey, J., Kuijpers, N., others, 1999. The COVEN Project: Exploring Applicative, Technical, and Usage Dimensions of Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments* 8(2), 218–236. <https://doi.org/https://doi.org/10.1162/105474699566189>
- Nukarinen, T., Kangas, J., Rantala, J., Koskinen, O., Raisamo, R., 2018. Evaluating Ray Casting and Two Gaze-Based Pointing Techniques for Object Selection in Virtual Reality, in: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology - VRST '18*. ACM, New York, NY, USA, S. 1–2. <https://doi.org/10.1145/3281505.3283382>
- O'Hagan, R.G., Zelinsky, A., Rougeaux, S., 2002. Visual Gesture Interfaces for Virtual Environments. *Interacting with Computers* 14(3), 231–250. [https://doi.org/10.1016/S0953-5438\(01\)00050-9](https://doi.org/10.1016/S0953-5438(01)00050-9)
- Oberdörfer, S., Fischbach, M., Latoschik, M.E., 2018. Effects of VE Transition Techniques on Presence, Illusion of Virtual Body Ownership, Efficiency, and Naturalness, in: *Proceedings of the Symposium on Spatial User Interaction - SUI '18*. ACM, New York, NY, USA, S. 89–99. <https://doi.org/10.1145/3267782.3267787>
- Oh, C.S., Bailenson, J.N., Welch, G.F., 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5, 1–35. <https://doi.org/10.3389/frobt.2018.00114>
- Ord, J.K., Getis, A., 1995. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application.

- Geographical Analysis 27(4), 286–306. <https://doi.org/10.1111/j.1538-4632.1995.tb00912.x>
- Pakkanen, T., Hakulinen, J., Jokela, T., Rakkolainen, I., Kangas, J., Piippo, P., Raisamo, R., Salmimaa, M., 2017. Interaction with WebVR 360° Video Player: Comparing Three Interaction Paradigms, in: 2017 IEEE Virtual Reality Conference - IEEEVR '17. IEEE, S. 279–280. <https://doi.org/10.1109/VR.2017.7892285>
- Passmore, P.J., Glancy, M., Philpot, A., Fields, B., 2017. 360 Cinematic Literacy: A Case Study, in: International Broadcasting Convention. S. 1–8.
- Pausch, R., Snoddy, J., Taylor, R., Watson, S., Haseltine, E., 1996. Disney's Aladdin: First Steps Toward Storytelling in Virtual Reality, in: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '96. ACM, New York, NY, USA, S. 193–203. <https://doi.org/10.1145/237170.237257>
- Pavel, A., Hartmann, B., Agrawala, M., 2017. Shot Orientation Controls for Interactive Cinematography with 360 Video, in: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST '17. ACM, New York, NY, USA, S. 289–297. <https://doi.org/10.1145/3126594.3126636>
- Peng, Y.-H., Tang, H.-H., Chen, M.Y., Hsi, M.-W., Taelle, P., Lin, T.-Y., Lai, P.-E., Hsu, L., Chen, T., Wu, T.-Y., Chen, Y.-A., 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18. ACM, New York, NY, USA, S. 1–10. <https://doi.org/10.1145/3173574.3173867>
- Penn P. Wu, 2017. Teaching College Classes with Virtual Reality. World Academy of Science, Engineering and Technology International Journal of Educational and Pedagogical Sciences 11(5), 1249–1254. <https://doi.org/doi.org/10.5281/zenodo.1130483>
- Perea, P., Morand, D., Nigay, L., 2017. [POSTER] Halo3D: A Technique for Visualizing Off-Screen Points of Interest in Mobile Augmented Reality, in: 2017 IEEE International Symposium on Mixed and Augmented Reality - ISMAR-Adjunct. IEEE, S. 170–175. <https://doi.org/10.1109/ISMAR-Adjunct.2017.58>
- Pfeuffer, K., Mayer, B., Mardanbegi, D., Gellersen, H., 2017. Gaze + Pinch Interaction in Virtual Reality, in: Proceedings of the 5th Symposium on Spatial User Interaction - SUI '17. ACM, New York, NY, USA, S. 99–108. <https://doi.org/10.1145/3131277.3132180>
- Philpot, A., Glancy, M., Passmore, P.J., Wood, A., Fields, B., 2017. User Experience of Panoramic Video in CAVE-like and Head Mounted Display Viewing Conditions, in: Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '17. ACM, New York, NY, USA, S. 65–75. <https://doi.org/10.1145/3077548.3077550>
- Piccolin, L., 2006. Rundumkinos. Hamburger Flimmern Heft 13 (2006) 22–27.
- Picinali, L., Afonso, A., Denis, M., Katz, B.F.G., 2014. Exploration of Architectural Spaces by Blind People Using Auditory Virtual Reality for the Construction of Spatial Knowledge. International Journal of Human Computer Studies 72(4), 393–407. <https://doi.org/10.1016/j.ijhcs.2013.12.008>
- Pillai, J.S., Ismail, A., Charles, H.P., 2017. Grammar of VR storytelling: Visual cues, in: ACM International Conference Proceeding Series. ACM, New York, NY, USA, S. 1–4. <https://doi.org/10.1145/3110292.3110300>
- Pope, V.C., Dawes, R., Schweiger, F., Sheikh, A., 2017. The Geometry of Storytelling: Theatrical Use of Space for 360-degree Videos and Virtual Reality, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17. ACM, New York, NY, USA, S. 4468–4478. <https://doi.org/10.1145/3025453.3025581>
- Posner, M.I., 1980. Orienting of attention. Quarterly Journal of Experimental Psychology 32(1), 3–25. <https://doi.org/10.1080/0033558008248231>

- Przybylski, A.K., Murayama, K., DeHaan, C.R., Gladwell, V., 2013. Motivational, Emotional, and Behavioral Correlates of Fear of Missing Out. *Computers in Human Behavior* 29(4), 1841–1848. <https://doi.org/10.1016/J.CHB.2013.02.014>
- Qian, Y.Y., Teather, R.J., 2017. The Eyes Don't Have It: An Empirical Comparison of Head-Based and Eye-Based Selection in Virtual Reality, in: *Proceedings of the 5th Symposium on Spatial User Interaction - SUI '17*. ACM, New York, NY, USA, S. 91–98. <https://doi.org/10.1145/3131277.3132182>
- Rai, Y., Gutiérrez, J., Le Callet, P., 2017. A Dataset of Head and Eye Movements for 360 Degree Images, in: *Proceedings of the 8th ACM on Multimedia Systems Conference - MMSys '17*. ACM, New York, NY, USA, S. 205–210. <https://doi.org/10.1145/3083187.3083218>
- Rantala, J., Kangas, J., Raisamo, R., 2017. Directional Cueing of Gaze with a Vibrotactile Headband, in: *Proceedings of the 8th Augmented Human International Conference on - AH '17*. ACM, New York, NY, USA, S. 1–7. <https://doi.org/10.1145/3041164.3041176>
- Razzaque, S., Kohn, Z., Whitton, M.C., 2001. Redirected Walking, in: *Eurographics 2001 - Short Presentations*. Eurographics Association, S. 1–6. <https://doi.org/10.2312/egs.20011036>
- Renner, P., Pfeiffer, T., 2017a. Evaluation of Attention Guiding Techniques for Augmented Reality-based Assistance in Picking and Assembly Tasks, in: *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion - IUI '17 Companion*. ACM, New York, NY, USA, S. 89–92. <https://doi.org/10.1145/3030024.3040987>
- Renner, P., Pfeiffer, T., 2017b. Attention Guiding Techniques Using Peripheral Vision and Eye Tracking for Feedback in Augmented-Reality-Based Assistance Systems, in: *2017 IEEE Symposium on 3D User Interfaces - 3DUI*. IEEE, S. 186–194. <https://doi.org/10.1109/3DUI.2017.7893338>
- Roth, C., 2015. Experiencing Interactive Storytelling. PhD Thesis, Amsterdam Vrije Universiteit.
- Rothe, S., Althammer, F., Khamis, M., 2018a. GazeRecall: Using Gaze Direction to Increase Recall of Details in Cinematic Virtual Reality, in: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia - MUM '18*. ACM, New York, NY, USA, S. 115–119. <https://doi.org/10.1145/3282894.3282903>
- Rothe, S., Brunner, H., Buschek, D., Hußmann, H., 2018b. Spaceline: A Way of Interaction in Cinematic Virtual Reality, in: *SUI 2018 - Proceedings of the Symposium on Spatial User Interaction*. ACM, New York, NY, USA, S. 179. <https://doi.org/10.1145/3267782.3274675>
- Rothe, S., Brunner, H., Hußmann, H., 2018c. Spaceline und Interaktion in Cinematischer Virtueller Realität, in: *Mensch und Computer 2018 - Tagungsband*. Gesellschaft für Informatik e.V., Bonn, S. 107–116. <https://doi.org/10.18420/muc2018-mci-0145>
- Rothe, S., Buschek, D., Hußmann, H., 2019a. Guidance in Cinematic Virtual Reality-Taxonomy, Research Status and Challenges. *Multimodal Technologies and Interaction* 3(1), 19–42. <https://doi.org/10.3390/mti3010019>
- Rothe, S., Hußmann, H., 2019. Spaceline: A Concept for Interaction in Cinematic Virtual Reality, in: *International Conference on Interactive Digital Storytelling - ICIDS 2019*. Springer, Cham, S. 115–119. https://doi.org/10.1007/978-3-030-33894-7_12
- Rothe, S., Hußmann, H., 2018. Guiding the Viewer in Cinematic Virtual Reality by Diegetic Cues, in: *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, Cham, S. 101–117. https://doi.org/10.1007/978-3-319-95270-3_7
- Rothe, S., Hußmann, H., Allary, M., 2017. Diegetic Cues for Guiding the Viewer in Cinematic Virtual Reality, in: *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17*. ACM, New York, NY, USA, S. 1–2. <https://doi.org/10.1145/3139131.3143421>
- Rothe, S., Kegeles, B., Allary, M., Hußmann, H., 2018d. The Impact of Camera Height in Cinematic Virtual

- Reality, in: Proceedings of the ACM Symposium on Virtual Reality Software and Technology - VRST '18. ACM, New York, NY, USA, S. 1–2. <https://doi.org/10.1145/3281505.3283383>
- Rothe, S., Kegeles, B., Hußmann, H., 2019b. Camera Heights in Cinematic Virtual Reality, in: Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '19. ACM, New York, NY, USA, S. 25–34. <https://doi.org/10.1145/3317697.3323362>
- Rothe, S., Montagud, M., Mai, C., Buschek, D., Hußmann, H., 2018e. Social Viewing in Cinematic Virtual Reality: Challenges and Opportunities, in: International Conference on Interactive Digital Storytelling - ICIDS 2018. Springer, Cham, S. 338–342. https://doi.org/10.1007/978-3-030-04028-4_37
- Rothe, S., Pothmann, P., Drewes, H., Hußmann, H., 2019c. Interaction Techniques for Cinematic Virtual Reality, in: IEEE Conference on Virtual Reality and 3D User Interfaces - IEEEVR '19. IEEE, S. 1733–1737. <https://doi.org/10.1109/VR.2019.8798189>
- Rothe, S., Sarakiotis, V., Hußmann, H., 2019d. Where to Place the Camera: Shot Sizes and Camera Distances in Cinematic Virtual Reality, in: Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology - VRST. ACM, New York, NY, USA, S. 1–2. <https://doi.org/10.1145/3359996.3364798>
- Rothe, S., Tran, K., Hußmann, H., 2018f. Dynamic Subtitles in Cinematic Virtual Reality, in: Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '18. ACM Press, New York, New York, USA, S. 209–214. <https://doi.org/https://doi.org/10.1145/3210825.3213556>
- Rothe, S., Tran, K., Hußmann, H., 2018g. Positioning of Subtitles in Cinematic Virtual Reality, in: International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments - ICAT-EGVE 2018. Eurographics Association, S. 1–8. <https://doi.org/10.2312/egve.20181307>
- Roussou, M., 2004. Learning by Doing and Learning Through Play: An Exploration of Interactivity in Virtual Environments for Children. *Computers in Entertainment* 2(1), 10–33. <https://doi.org/10.1145/973801.973818>
- Rovamo, J., Raninen, A., 1988. Critical Flicker Frequency as a Function of Stimulus Area and Luminance at Various Eccentricities in Human Cone Vision: A Revision of Granit-Harper and Ferry-Porter Laws. *Vision Research* 28(7), 785–790. [https://doi.org/10.1016/0042-6989\(88\)90025-9](https://doi.org/10.1016/0042-6989(88)90025-9)
- Rovelo, G., Vanacken, D., Luyten, K., Abad, F., Camahort, E., 2014. Multi-Viewer Gesture-Based Interaction for Omni-Directional Video, in: Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14. ACM Press, New York, New York, USA, S. 4077–4086. <https://doi.org/10.1145/2556288.2557113>
- Sargunam, S.P., Moghadam, K.R., Suhail, M., Ragan, E.D., 2017. Guided Head Rotation and Amplified Head Rotation: Evaluating Semi-natural Travel and Viewing Techniques in Virtual Reality, in: 2017 IEEE Virtual Reality Conference - IEEEVR '17. IEEE, S. 19–28. <https://doi.org/10.1109/VR.2017.7892227>
- Sassatelli, L., Pinna-Déry, A.-M., Winckler, M., Dambra, S., Samela, G., Pighetti, R., Aparicio-Pardo, R., 2018. Snap-changes: A Dynamic Editing Strategy for Directing Viewer's Attention in Streaming Virtual Reality Videos, in: Proceedings of the 2018 International Conference on Advanced Visual Interfaces - AVI '18. ACM, New York, NY, USA, S. 1–5. <https://doi.org/10.1145/3206505.3206553>
- Schultze, U., Brooks, J.A.M., 2019. An Interactional View of Social Presence: Making the Virtual Other “Real”. *Information Systems Journal* 29(3), 707–737. <https://doi.org/10.1111/isj.12230>
- Sekuler, A.B., Bennett, P.J., 2000. Effects of Aging on the Useful Field of View. *Experimental Aging Research* 26(2), 103–120. <https://doi.org/10.1080/036107300243588>
- Serrano, A., Sitzmann, V., Ruiz-Borau, J., Wetzstein, G., Gutierrez, D., Masia, B., 2017. Movie Editing and

- Cognitive Event Segmentation in Virtual Reality Video. *ACM Transactions on Graphics* 36(4), 1–12. <https://doi.org/10.1145/3072959.3073668>
- Setti, F., Russell, C., Bassetti, C., Cristani, M., 2014. F-formation Detection: Individuating Free-standing Conversational Groups in Images. *PLOS ONE* 10(5), e0123783. <https://doi.org/10.1371/journal.pone.0123783>
- Shah, P., Miyake, A., 2005. *The Cambridge Handbook of Visuospatial Thinkin*. Cambridge University Press.
- Sheikh, A., Brown, A., Evans, M., Watson, Z., 2016. Directing Attention in 360-Degree Video, in: *IBC 2016 Conference*. Institution of Engineering and Technology, Amsterdam, Netherlands, S. 1–9. <https://doi.org/10.1049/ibc.2016.0029>
- Shin, D.-H., Kim, J., 2015. Social Viewing Behavior in Social TV: Proposing a New Concept of Socio-Usability. *Online Information Review* 39(3), 416–434. <https://doi.org/10.1108/OIR-12-2014-0299>
- Silva, A., Raimundo, G., Paiva, A., 2003. Tell Me That Bit Again... Bringing Interactivity to a Virtual Storyteller, in: *International Conference on Virtual Storytelling*. Springer, Berlin, Heidelberg, S. 146–154. https://doi.org/10.1007/978-3-540-40014-1_19
- Sinai, M.J., Ooi, T.L., He, Z.J., 1998. Terrain Influences the Accurate Judgement of Distance. *Nature* 395(6701), 497–500. <https://doi.org/10.1038/26747>
- Sitzmann, V., Serrano, A., Pavel, A., Agrawala, M., Gutierrez, D., Masia, B., Wetzstein, G., 2018. Saliency in VR: How Do People Explore Virtual Environments? *IEEE Transactions on Visualization and Computer Graphics* 24(4), 1633–1642. <https://doi.org/10.1109/TVCG.2018.2793599>
- Skarbez, R., Brooks, Jr., F.P., Whitton, M.C., 2017. A Survey of Presence and Related Concepts. *ACM Computing Surveys* 50(6), 1–39. <https://doi.org/10.1145/3134301>
- Slater, M., Sanchez-Vives, M. V., 2016. Enhancing Our Lives with Immersive Virtual Reality. *Frontiers in Robotics and AI* 3, 74. <https://doi.org/10.3389/frobt.2016.00074>
- Smith, T.J., 2013. Watching You Watch Movies: Using Eye Tracking to Inform Cognitive Film Theory, in: *Psychocinematics*. Oxford University Press, S. 165–192. <https://doi.org/10.1093/acprof:oso/9780199862139.003.0009>
- Smith, T.J., 2012. The Attentional Theory of Cinematic Continuity. *Projections* 6(1), 1–27. <https://doi.org/10.3167/proj.2012.060102>
- Smith, T.J., 2006. *An Attentional Theory of Continuity Editing*. PhD Thesis, University of Edinburgh. College of Science and Engineering. School of Informatics.
- Smith, T.J., Henderson, M.J., 2008. Edit Blindness: The Relationship between Attention and Global Change Blindness in Dynamic Scenes. *Journal of Eye Movement Research* 2(2), 1–17. <https://doi.org/https://doi.org/10.16910/jemr.2.2.6>
- Smith, T.J., Martin-Portugues Santacreu, J.Y., 2017. Match-Action: The Role of Motion and Audio in Creating Global Change Blindness in Film. *Media Psychology* 20(2), 317–348. <https://doi.org/10.1080/15213269.2016.1160789>
- Smith, W.S., Tadmor, Y., 2013. Nonblurred Regions Show Priority for Gaze Direction over Spatial Blur. *Quarterly Journal of Experimental Psychology* 66(5), 927–945. <https://doi.org/10.1080/17470218.2012.722659>
- Snyder, J.P., 1987. *Map Projections: A Working Manual*. US Government Printing Office, Washington, DC.
- Soler-Adillon, J., Sora, C., 2018. Immersive Journalism and Virtual Reality, in: *Interaction in Digital News Media*. Springer International Publishing, Cham, S. 55–83. https://doi.org/10.1007/978-3-319-96253-5_4

- Songchitruksa, P., Zeng, X., 2010. Getis–Ord Spatial Statistics to Identify Hot Spots by Using Incident Management Data. *Transportation Research Record: Journal of the Transportation Research Board* 2165(1), 42–51. <https://doi.org/10.3141/2165-05>
- Souriau, E., 1951. La structure de l'univers filmique et le vocabulaire de la filmologie. | Interdisciplinary Center for Narratology, in: *Revue internationale de filmologie*. 7-8. S. 231–240.
- Špakov, O., Majaranta, P., 2012. Enhanced Gaze Interaction Using Simple Head Gestures, in: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. ACM, New York, NY, USA, S. 705–710. <https://doi.org/10.1145/2370216.2370369>
- Steed, A., 2006. Towards a General Model for Selection in Virtual Environments, in: *3D User Interfaces - 3DUI '06*. IEEE, S. 103–110. <https://doi.org/10.1109/VR.2006.134>
- Steinicke, F., 2018. Die Zukunft der erweiterten Realität, in: *Recht & Netz*. Nomos Verlagsgesellschaft mbH & Co. KG, S. 457–468. <https://doi.org/10.5771/9783845293288-457>
- Steinicke, F., Bruder, G., Jerald, J., Frenz, H., Lappe, M., 2010. Estimation of Detection Thresholds for Redirected Walking Techniques. *IEEE Transactions on Visualization and Computer Graphics* 16(1), 17–27. <https://doi.org/10.1109/TVCG.2009.62>
- Steinicke, F., Bruder, G., Jerald, J., Frenz, H., Lappe, M., 2008. Analyses of Human Sensitivity to Redirected Walking, in: *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology - VRST '08*. ACM, New York, NY, USA, S. 149–156. <https://doi.org/10.1145/1450579.1450611>
- Steinman, R.M., Pizlo, Z., Pizlo, F.J., 2000. Phi is Not Beta, and Why Wertheimer's Discovery Launched the Gestalt Revolution. *Vision Research* 40(17), 2257–2264. [https://doi.org/10.1016/S0042-6989\(00\)00086-9](https://doi.org/10.1016/S0042-6989(00)00086-9)
- Stojšić, I., Ivkov-Džigurski, A., Maričić, O., 2019. Virtual Reality as a Learning Tool: How and Where to Start with Immersive Teaching, in: *Didactics of Smart Pedagogy*. Springer International Publishing, Cham, S. 353–369. https://doi.org/10.1007/978-3-030-01551-0_18
- Stratmann, T.C., Löcken, A., Gruenefeld, U., Heuten, W., Boll, S., 2018. Exploring Vibrotactile and Peripheral Cues for Spatial Attention Guidance, in: *Proceedings of the 7th ACM International Symposium on Pervasive Displays - PerDis '18*. ACM, New York, NY, USA, S. 1–8. <https://doi.org/10.1145/3205873.3205874>
- Subramanian, R., Shankar, D., Sebe, N., Melcher, D., 2014. Emotion Modulates Eye Movement Patterns and Subsequent Memory for the Gist and Details of Movie Scenes. *Journal of Vision* 14(3), 31–31. <https://doi.org/10.1167/14.3.31>
- Suh, A., Wang, G., Gu, W., Wagner, C., 2018. Enhancing Audience Engagement Through Immersive 360-Degree Videos: An Experimental Study, in: *International Conference on Augmented Cognition*. Springer, Cham, S. 425–443. https://doi.org/10.1007/978-3-319-91470-1_34
- Sukan, M., Elvezio, C., Oda, O., Feiner, S., Tversky, B., 2014. ParaFrustum: Visualization Techniques for Guiding a User to a Constrained Set of Viewing Positions and Orientations, in: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology - UIST '14*. ACM, New York, NY, USA, S. 331–340. <https://doi.org/10.1145/2642918.2647417>
- Suma, E.A., Bruder, G., Steinicke, F., Krum, D.M., Bolas, M., 2012. A Taxonomy for Deploying Redirection Techniques in Immersive Virtual Environments, in: *2012 IEEE Virtual Reality Conference - IEEEVR '12*. IEEE, S. 43–46. <https://doi.org/10.1109/VR.2012.6180877>
- Swenberg, T., Eriksson, P.E., 2018. Effects of Continuity or Discontinuity in Actual Film Editing. *Empirical Studies of the Arts* 36(2), 222–246. <https://doi.org/10.1177/0276237417744590>
- Sylvester, T., 2013. *Designing Games: A Guide to Engineering Experiences*. O'Reilly Media, Inc.

- Syrett, H., Calvi, L., van Gisbergen, M., 2016. The Oculus Rift Film Experience: A Case Study on Understanding Films in a Head Mounted Display, in: *International Conference on Intelligent Technologies for Interactive Entertainment*. Springer, Cham, S. 197–208. https://doi.org/10.1007/978-3-319-49616-0_19
- Tanak, N., 2015. *Interactive Cinema*. MediaLAB Amsterdam.
- Tang, A., Fakourfar, O., 2017. Watching 360° Videos Together, in: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM, New York, NY, USA, S. 4501–4506. <https://doi.org/10.1145/3025453.3025519>
- Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., Nießner, M., 2016. FaceVR: Real-Time Facial Reenactment and Eye Gaze Control in Virtual Reality. *ACM Transactions on Graphics* 37(2), 1–15. <https://doi.org/10.1145/3182644>
- Thompson, K., Bordwell, D., 2010. *Film History : An Introduction*. McGraw-Hill Higher Education.
- Tikka, P., 2004. (Interactive) Cinema as a Model of Mind. *Digital Creativity* 15(1), 14–17. <https://doi.org/10.1076/digc.15.1.14.28151>
- Treisman, A.M., Gelade, G., 1980. A Feature-Integration Theory of Attention. *Cognitive Psychology* 12(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Tyler, C.W., Hamer, R.D., 1993. Eccentricity and the Ferry–Porter Law. *Journal of the Optical Society of America A* 10(9), 2084–2087. <https://doi.org/10.1364/JOSAA.10.002084>
- Upenik, E., Ebrahimi, T., 2017. A Simple Method to Obtain Visual Attention Data in Head Mounted Virtual Reality, in: *2017 IEEE International Conference on Multimedia & Expo Workshops - ICMEW*. IEEE, S. 73–78. <https://doi.org/10.1109/ICMEW.2017.8026231>
- Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., Theeuwes, J., 2008. Pip and Pop: Nonspatial Auditory Signals Improve Spatial Visual Search. *Journal of Experimental Psychology: Human Perception and Performance* 34(5), 1053–1065. <https://doi.org/10.1037/0096-1523.34.5.1053>
- Veas, E.E., Mendez, E., Feiner, S.K., Schmalstieg, D., 2011. Directing Attention and Influencing Memory with Visual Saliency Modulation, in: *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*. ACM, New York, NY, USA, S. 1471–1480. <https://doi.org/10.1145/1978942.1979158>
- Velichkovsky, B., Sprenger, A., Unema, P., 1997. Towards Gaze-Mediated Interaction: Collecting Solutions of the “Midas Touch Problem”, in: *Human-Computer Interaction - INTERACT '97*. Springer US, Boston, MA, S. 509–516. https://doi.org/10.1007/978-0-387-35175-9_77
- Verdugo, R., Nussbaum, M., Corro, P., Nuñez, P., Navarrete, P., 2011. Interactive Films and Coconstruction. *ACM Transactions on Multimedia Computing, Communications, and Applications* 7(4), 1–24. <https://doi.org/10.1145/2043612.2043617>
- Vig, E., Dorr, M., Barth, E., 2011. Learned Saliency Transformations for Gaze Guidance, in: *Proceedings SPIE 7865, Human Vision and Electronic Imaging XVI*, 78650W. SPIE, S. 311–321. <https://doi.org/10.1117/12.876377>
- Voorveld, H.A.M., Viswanathan, V., 2015. An Observational Study on How Situational Factors Influence Media Multitasking With TV: The Role of Genres, Dayparts, and Social Viewing. *Media Psychology* 18(4), 499–526. <https://doi.org/10.1080/15213269.2013.872038>
- Vosmeer, M., Schouten, B., 2014. Interactive Cinema: Engagement and Interaction, in: *International Conference on Interactive Digital Storytelling*. Springer, Cham, S. 140–147. https://doi.org/10.1007/978-3-319-12337-0_14
- Vy, Q. V., Fels, D.I., 2010. Using Placement and Name for Speaker Identification in Captioning, in: *Computers Helping People with Special Needs*. Springer, Berlin, Heidelberg, S. 247–254.

- https://doi.org/10.1007/978-3-642-14097-6_40
- Vy, Q. V., Fels, D.I., 2009. Using Avatars for Improving Speaker Identification in Captioning, in: IFIP Conference on Human-Computer Interaction. Springer, Berlin, Heidelberg, S. 916–919. https://doi.org/10.1007/978-3-642-03658-3_110
- Wakefield, G., Hollerer, T., Kuchera-Morin, J., Roberts, C., Wright, M., 2013. Spatial Interaction in a Multiuser Immersive Instrument. *IEEE Computer Graphics and Applications* 33(6), 14–20. <https://doi.org/10.1109/MCG.2013.99>
- Waldin, N., Waldner, M., Viola, I., 2017. Flicker Observer Effect: Guiding Attention Through High Frequency Flicker in Images. *Computer Graphics Forum* 36(2), 467–476. <https://doi.org/10.1111/cgf.13141>
- Wang, G., Gu, W., Suh, A., 2018. The Effects of 360-Degree VR Videos on Audience Engagement: Evidence from the New York Times, in: International Conference on HCI in Business, Government, and Organizations - HCIBGO. Springer, Cham, S. 217–235. https://doi.org/10.1007/978-3-319-91716-0_17
- Weisz, J.D., Kiesler, S., Zhang, H., Ren, Y., Kraut, R.E., Konstan, J.A., 2007. Watching Together: Integrating Text Chat with Video, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07. ACM, New York, NY, USA, S. 877–886. <https://doi.org/10.1145/1240624.1240756>
- Wertheimer, M., 1938. Gestalt Theory, in: Ellis, W.D. (Hrsg.), *Source Book of Gestalt Psychology*. Kegan Paul, Trench, Trubner & Company. <https://doi.org/10.1037/11496-001>
- Wilcox, L.M., Allison, R.S., Elfassy, S., Grelik, C., 2006. Personal Space in Virtual Reality. *ACM Transactions on Applied Perception* 3(4), 412–428. <https://doi.org/10.1145/1190036.1190041>
- Wilson, P.N., Foreman, N., Stanton, D., 1997. Virtual Reality, Disability and Rehabilitation. *Disability and Rehabilitation* 19(6), 213–220. <https://doi.org/10.3109/09638289709166530>
- Witmer, B.G., Singer, M.J., 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7(3), 225–240. <https://doi.org/10.1162/105474698565686>
- Wolfe, J.M., Horowitz, T.S., 2017. Five Factors that Guide Attention in Visual Search. *Nature Human Behaviour* 1(3), 1–8. <https://doi.org/10.1038/s41562-017-0058>
- Wolfe, J.M., Horowitz, T.S., 2004. What Attributes Guide the Deployment of Visual Attention and How Do They Do It? *Nature Reviews Neuroscience* 5(6), 495–501. <https://doi.org/10.1038/nrn1411>
- Wraga, Maryjane, 1999. The Role of Eye Height in Perceiving Affordances and Object Dimensions. *Perception & Psychophysics* 61(3), 490–507. <https://doi.org/10.3758/BF03211968>
- Wraga, M., 1999. Using Eye Height in Different Postures to Scale the Heights of Objects. *Journal of Experimental Psychology. Human Perception and Performance* 25(2), 518–530. <https://doi.org/10.1037//0096-1523.25.2.518>
- Wright, R.D., 1998. *Visual attention*. Oxford University Press.
- Yarbus, A.L., 1967. *Eye Movements and Vision*. Springer US, Boston, MA. <https://doi.org/10.1007/978-1-4899-5379-7>
- Yee, N., Bailenson, J.N., Ducheneaut, N., 2009. The Proteus Effect. *Communication Research* 36(2), 285–312. <https://doi.org/10.1177/0093650208330254>
- Yeh, M., Wickens, C.D., Seagull, F.J., 1999. Target Cuing in Visual Search: The Effects of Conformality and Display Location on the Allocation of Visual Attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 41(4), 524–542. <https://doi.org/10.1518/001872099779656752>

- Yu, C., Gu, Y., Yang, Z., Yi, X., Luo, H., Shi, Y., 2017. Tap, Dwell or Gesture?: Exploring Head --Based Text Entry Techniques for HMDs, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17. ACM, New York, NY, USA, S. 4479–4488. <https://doi.org/10.1145/3025453.3025964>
- Zacks, J.M., 2010. How We Organize Our Experience into Events. *Psychological Science Agenda* 24, 1–8.
- Zacks, J.M., Speer, N.K., Swallow, K.M., Braver, T.S., Reynolds, J.R., 2007. Event Perception: A Mind/Brain Perspective. *Psychological Bulletin* 133(2), 273–293. <https://doi.org/10.1037/0033-2909.133.2.273>
- Zellweger, P.T., Mackinlay, J.D., Good, L., Stefik, M., Baudisch, P., 2003. City Lights: Contextual Views in Minimal Space, in: CHI '03 Extended Abstracts on Human Factors in Computing Systems - CHI '03. ACM, New York, NY, USA, S. 838–839. <https://doi.org/10.1145/765891.766022>
- Zielasko, D., Freitag, S., Rausch, D., Law, Y.C., Weyers, B., Kuhlen, T.W., 2015. BlowClick: A Non-Verbal Vocal Input Metaphor for Clicking, in: Proceedings of the 3rd ACM Symposium on Spatial User Interaction - SUI '15. ACM, New York, NY, USA, S. 20–23. <https://doi.org/10.1145/2788940.2788953>

Webseitenverzeichnis

- Amann, C., 2012. Kreative Geographie - Lexikon der Filmbegriffe. <https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=5685> (zugegriffen 8.Feb.2020).
- ARD, 2018. Barrierefreie Angebote der ARD. http://www.ard.de/home/die-ard/fakten/Barrierefreie_Angebote_der_ARD/4036078/index.html (zugegriffen 19.Feb.2020).
- Athena Studios, 2019. Athena Studios - Adobe Character Animator. https://www.athenastudios.com/project_adobecharacter.asp (zugegriffen 28.Jan.2020).
- Beller, H., Wulff, H.J., 2012. 30-Grad-Regel - Lexikon der Filmbegriffe. <https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=3425> (zugegriffen 8.Feb.2020).
- Bendel, O., 2018. Virtuelle Realität. Gabler Wirtschaftslexikon. <https://wirtschaftslexikon.gabler.de/definition/virtuelle-realitaet-54243/version-277293> (zugegriffen 9.Apr.2020).
- Bender, T., Wulff, H.J., Denzer, K., 2012. Einstellung - Lexikon der Filmbegriffe. <https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=135> (zugegriffen 8.Feb.2020).
- Bloom, P., 2017b. It's Ridiculous to Use Virtual Reality to Empathize With Refugees - The Atlantic. <https://www.theatlantic.com/technology/archive/2017/02/virtual-reality-wont-make-you-more-empathetic/515511/> (zugegriffen 28.Dez.2018).
- Brillhart, J., 2016. In the Blink of a Mind — Attention. <https://medium.com/the-language-of-vr/in-the-blink-of-a-mind-attention-1fdff60fa045> (zugegriffen 10.Okt.2019).
- British Library, 2017. Section of the Rotunda, Leicester Square. <https://www.bl.uk/collection-items/section-of-the-rotunda-leicester-square#> (zugegriffen 9.März.2020).
- Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., Glancy, M., 2018. Exploring Subtitle Behaviour for 360° Video. <https://www.bbc.co.uk/rd/publications/whitepaper330> (zugegriffen 12.Nov.2018).
- Byford, S., 2012. Kogeto iCONIC Panoramic Video Coming to Galaxy Nexus. <https://www.theverge.com/2012/1/6/2687245/kogeto-iconic-panoramic-lens-galaxy-nexus>

- (zugegriffen 9.März.2020).
- Cambridge Dictionary, 2019. Interaction.
<https://dictionary.cambridge.org/de/worterbuch/englisch/interaction> (zugegriffen 30.März.2020).
- Crew United, 2019. Bauhaus in Bayern. https://www.crew-united.com/de/Bauhaus-in-Bayern-360-VR__231032.html (zugegriffen 4.Apr.2020).
- Dickey, J., 2014. Virtual Reality is Exploding at this Year's Sundance Film Festival.
<https://mashable.com/2014/12/15/virtual-reality-sundance-film-festival> (zugegriffen 9.März.2020).
- esri, 2018. Visualize Space Time Cube in 3D. <http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/visualizecube3d.htm> (zugegriffen 5.Juli.2018).
- facebook, 2016. New Publisher Tools for 360 Video.
<https://www.facebook.com/facebookmedia/blog/new-publisher-tools-for-360-video> (zugegriffen 17.Feb.2020).
- Ford, L., 2017. Virtual Reality, 19th Century Style: The History of the Panorama and Balloon View - OpenLearn - Open University. <https://www.open.edu/openlearn/history-the-arts/visual-art/virtual-reality-19th-century-style-the-history-the-panorama-and-balloon-view> (zugegriffen 9.März.2020).
- Headjack, 2018. Create Virtual Reality Video Apps in Minutes instead of Months. <https://headjack.io/> (zugegriffen 30.Juni.2018).
- Hinderks, A., Schrepp, M., Thomaschewski, J., 2018. User Experience Questionnaire (UEQ).
<https://www.ueq-online.org/> (zugegriffen 7.Juli.2019).
- Hünigen, J., 2012a. Schnitt 2 - Lexikon der Filmbegriffe. <https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=4437> (zugegriffen 8.Feb.2020).
- Hünigen, J., 2012b. Schnitt 1 - Lexikon der Filmbegriffe. <https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=4436> (zugegriffen 8.Feb.2020).
- Hunt, C., 2016. Field of View Face-off: Rift vs Vive vs Gear VR vs PSVR. <https://www.vrheads.com/field-view-faceoff-rift-vs-vive-vs-gear-vr-vs-psvr> (zugegriffen 13.Juli.2019).
- Immersive Accessibility, 2018. ImAc. <https://www.imac-project.eu/> (zugegriffen 6.Apr.2020).
- InstaVR, 2018. Make Your VR Apps in Minutes. <http://www.instavr.co/> (zugegriffen 30.Juni.2018).
- iTechWearer, 2017. The Secret to the Best Home Theater: FOV – Field of View Technology.
<https://www.itechwearer.com/the-secret-to-the-best-home-theater-fov-cinera/> (zugegriffen 19.Juli.2019).
- Jaunt, 2017. The Cinematic VR Field Guide. <https://www.jauntvr.com/cdn/uploads/jaunt-vr-field-guide.pdf> (zugegriffen 20.Mai.2018).
- Kamps, U., 2018a. Datenanalyse. Gabler Wirtschaftslexikon.
<https://wirtschaftslexikon.gabler.de/definition/datenanalyse-30331/version-253916> (zugegriffen 9.Apr.2020).
- Kamps, U., 2018b. Deskriptive Statistik. Gabler Wirtschaftslexikon.
<https://wirtschaftslexikon.gabler.de/definition/deskriptive-statistik-31622/version-255175> (zugegriffen 7.Apr.2020).
- Kamps, U., 2018c. Inferenzstatistik. Gabler Wirtschaftslexikon.
<https://wirtschaftslexikon.gabler.de/definition/inferenzstatistik-40284/version-263672> (zugegriffen 7.Apr.2020).

- Kaur, G., 2018. Virtual Reality. <https://medium.com/@kaurgagan073/virtual-reality-5a0164a2a3c2> (zugegriffen 9.März.2020).
- Kent, T., 2015. An Ethical Reality Check for Virtual Reality Journalism. <https://medium.com/@tjrkent/an-ethical-reality-check-for-virtual-reality-journalism-8e5230673507> (zugegriffen 29.Dez.2018).
- Kirchlechner, J., 2013. Kinematografie. Berufsverband Kinematografie - German Society of Cinematographers. <https://kinematografie.org/berufsverband/kinematografie.php> (zugegriffen 6.Apr.2020).
- Lomography, 2017. History · Spinner 360°. <https://microsites.lomography.com/spinner-360/history/> (zugegriffen 9.März.2020).
- Matroska, 2020. What is Matroska? <https://www.matroska.org/technical/whatis/index.html> (zugegriffen 3.Apr.2020).
- Matroska, 2018. Matroska Media Container. <https://www.matroska.org/> (zugegriffen 3.Apr.2020).
- Matroska, 2015. libEBML. <http://matroska-org.github.io/libebml/> (zugegriffen 3.Apr.2020).
- media+art+innovation, 2014. Joseph Puchberger: Panoramic Camera 1843. <https://mediartinnovation.com/2014/05/21/joseph-puchberger-panoramic-camera-1843/> (zugegriffen 9.März.2020).
- Medieninnovationszentrum Babelsberg, 2018. Constructive VR. <https://www.miz-babelsberg.de/constructive-vr> (zugegriffen 28.Dez.2018).
- Milk, C., 2015. How Virtual Reality Can Create the Ultimate Empathy Machine. <https://en.unesco.org/interculturaldialogue/resources/276> (zugegriffen 28.Dez.2018).
- Mücher, M., 2018. AAF - Definition im Fachwörterbuch / Lexikon - BET. <https://www.bet.de/lexikon/aaf/> (zugegriffen 3.Apr.2020).
- Mücher, M., 2016. Quicktime - Definition im Fachwörterbuch / Lexikon - BET. <https://www.bet.de/lexikon/quicktime/> (zugegriffen 3.Apr.2020).
- Mücher, M., 2014a. AVI - Definition im Fachwörterbuch / Lexikon - BET. <https://www.bet.de/lexikon/avi/> (zugegriffen 3.Apr.2020).
- Mücher, M., 2014b. MXF - Definition im Fachwörterbuch / Lexikon - BET. <https://www.bet.de/lexikon/mxf/> (zugegriffen 2.Apr.2020).
- Mücher, M., 2013. MPEG-4/Part 14 - Definition im Fachwörterbuch / Lexikon - BET. <https://www.bet.de/lexikon/mpeg4part14/> (zugegriffen 3.Apr.2020).
- Mücher, M., 2012. OMF - Definition im Fachwörterbuch / Lexikon - BET. <https://www.bet.de/lexikon/omf/> (zugegriffen 3.Apr.2020).
- National Center for Health Statistics, 2017. FastStats - Body Measurements. <https://www.cdc.gov/nchs/fastats/body-measurements.htm> (zugegriffen 5.Aug.2018).
- Niebur, E., 2007. Saliency Map. Scholarpedia. <https://doi.org/10.4249/scholarpedia.2675>
- Oculus, 2018. Space Explorers für Oculus Rift. <https://www.oculus.com/experiences/rift/1683069235108094/> (zugegriffen 4.Apr.2020).
- Panotour, 2018. Kolor Panotour - Virtual Tour Creation Software. <http://www.kolor.com/panotour/> (zugegriffen 30.Juni.2018).
- Piccolin, L., 2012. Circarama - Lexikon der Filmbegriffe. <https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=110> (zugegriffen 9.März.2020).
- Point in Time Studios, 2018. The History of Virtual Reality. <https://pointintimestudios.com/the-history->

- of-virtual-reality/ (zugegriffen 9.März.2020).
- Reismann, S., 2020. Sensorama – Virtual Reality in 1962. <https://www.netzpiloten.de/sensorama-virtual-reality-in-1962/> (zugegriffen 9.März.2020).
- Rodriguez, F., 2017. The world as you see it with VR180. YouTube. <https://youtube.googleblog.com/2017/06/the-world-as-you-see-it-with-vr180.html> (zugegriffen 3.Aug.2019).
- Rouse, M., 2017. Filed field View (FOV). <https://whatis.techtarget.com/definition/field-of-view-FOV> (zugegriffen 9.Apr.2020).
- Samsung, 2018. Samsung Gear VR with Controller - The Official Samsung Galaxy Site. <https://www.samsung.com/global/galaxy/gear-vr/> (zugegriffen 20.Sep.2019).
- Samsung, 2016. Samsung Gear VR. <https://www.samsung.com/at/wearables/gear-vr-r323/> (zugegriffen 8.Apr.2020).
- Schubert, T., Friedmann, F., Regenbrecht, H., 2002. igroup Presence Questionnaire (IPQ). <http://www.igroup.org/pq/ipq/index.php> (zugegriffen 30.Juni.2018).
- Sergej Eisenstein, 1929. Tagebuch 5.August 1929. <https://www.fondation-langlois.org/html/e/page.php?NumPage=749> (zugegriffen 19.Feb.2020).
- Slater, M., 2003. A Note on Presence Terminology. http://www0.cs.ucl.ac.uk/research/vr/Projects/Presencia/ConsortiumPublications/ucl_cs_papers/presence-terminology.htm (zugegriffen 10.Okt.2019).
- SRF, 2018. Echorama - Ernst A. Heiniger. Schweizer Radio und Fernsehen SRF, Zürich. <https://www.concept360.ch/de/srf-echorama-ernst-a-heiniger/> (zugegriffen 9.März.2020).
- Stanford Studie, 2017. Becoming Homeless: A Human Experience Official Trailer - YouTube. https://www.youtube.com/watch?time_continue=36&v=OK2yJlCahDE (zugegriffen 30.Dez.2018).
- Toczyński, P., 2018. Editing with an Eye-trace in Mind: Is the Rule of Six Incorrect? <https://nofilmschool.com/2018/08/editing-eye-trace-mind-rule-six-incorrect> (zugegriffen 2.Feb.2020).
- Unity, 2017. GitHub - Unity-Technologies/360-Video-Heatmaps: Heatmaps for 360 Video. <https://github.com/Unity-Technologies/360-Video-Heatmaps> (zugegriffen 17.Feb.2020).
- Valve Corporation, 2017. Field of View - Deep Dive - Valve Index - Upgrade Your Experience. <https://www.valvesoftware.com/en/index/deep-dive/fov> (zugegriffen 17.Juli.2019).
- Viar360, 2016. Vitrual Reality Analytics: Why Do You Need It and How To Use It? <https://www.viar360.com/virtual-reality-analytics/> (zugegriffen 17.Feb.2020).
- Viond, 2018. Create and Publish Your Own Interactive 360 VR Experiences. <https://www.viond.io/> (zugegriffen 30.Juni.2018).
- Ward, L., 2008. Attention. Scholarpedia. <https://doi.org/10.4249/scholarpedia.1538>
- WebM-Consortium, 2019. The WebM Project. <https://www.webmproject.org/> (zugegriffen 3.Apr.2020).
- Whitwam, R., 2018. Google Opens Up VR180 Standard for Virtual Reality Photos and Videos - ExtremeTech. <https://www.extremetech.com/mobile/267271-google-opens-up-vr180-standard-for-virtual-reality-photos-and-videos> (zugegriffen 5.Aug.2019).
- Wikipedia, 2020. Cinéorama. <https://en.wikipedia.org/w/index.php?title=Cinéorama&oldid=938983191> (zugegriffen 9.März.2020).
- Wikipedia, 2018. Sichtfeld. <https://de.wikipedia.org/w/index.php?title=Sichtfeld&oldid=181507432> (zugegriffen 1.Apr.2020).

- Wistia, 2017. Heatmaps. <https://wistia.com/support/stats/heatmaps> (zugegriffen 17.Feb.2020).
- Wonda VR, 2019. Create and Share 360° Interactive VR Experience. <http://www.wondavr.com/> (zugegriffen 30.Juni.2018).
- Wulff, H.J., 2014. Montage - Lexikon der Filmbegriffe. <https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=1257> (zugegriffen 8.Feb.2020).
- Wulff, H.J., 2012a. Schärfeverlagerung. Lexikon der Filmbegriffe. <http://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=1876> (zugegriffen 9.Apr.2020).
- Wulff, H.J., 2012b. Achsensprung - Lexikon der Filmbegriffe. <https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=1342> (zugegriffen 6.März.2020).
- Wulff, H.J., 2012c. Mis-en-scène - Lexikon der Filmbegriffe. <http://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=4741> (zugegriffen 4.Feb.2020).
- Wulff, H.J., 2012d. Kuleschow-Effekt - Lexikon der Filmbegriffe. <https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=238> (zugegriffen 8.Feb.2020).
- Wulff, H.J., 2012e. Szene - Lexikon der Filmbegriffe. <https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=1501> (zugegriffen 8.Feb.2020).
- Wulff, H.J., 2012f. Transition - Lexikon der Filmbegriffe. <http://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=5584> (zugegriffen 11.März.2020).
- YouTube, 2017. YouTube Creator Blog: Hot and Cold: Heatmaps in VR. <https://youtube-creators.googleblog.com/2017/06/hot-and-cold-heatmaps-in-vr.html> (zugegriffen 17.Feb.2020).

Filmverzeichnis

- Aardman Animations, 2015. Special Delivery. YouTube. <https://www.youtube.com/watch?v=j6UzWLumnnw> (zugegriffen 13.Apr.2020).
- Arora, G., Milk, C., 2015. Clouds Over Sidra. WITHIN. <https://www.with.in/watch/clouds-over-sidra/> (zugegriffen 29.Dez.2018).
- Bigelow, K., 2017. The Protectors - Walk in the Ranger's Shoes. YouTube. <https://www.youtube.com/watch?v=RuGeeGRdYIQ> (zugegriffen 28.Dez.2018).
- de la Peña, N., 2014. Project Syria. YouTube. <https://www.youtube.com/watch?v=HtZrSb84JPE> (zugegriffen 29.Dez.2018).
- de la Peña, N., 2012. Hunger in Los Angeles. YouTube. https://www.youtube.com/watch?v=wwXPP_OOfzc (zugegriffen 29.Dez.2018).
- de la Peña, N., 2007. Gone Gitmo. YouTube. <https://www.youtube.com/watch?v=UHTLqqALPv0> (zugegriffen 29.Dez.2018).
- Felix & Paul Studios, 2017. The People's House - Inside the White House with Barack and Michelle Obama. YouTube. <https://www.youtube.com/watch?v=bqW2qm02jwl> (zugegriffen 28.Dez.2018).
- Felix & Paul Studios, 2016. Nomads. Oculus. https://www.oculus.com/experiences/gear-vr/820253268102944/?locale=de_DE (zugegriffen 30.März.2020).
- Google Research and Daydream Labs, 2017. Headset Removal. YouTube. https://www.youtube.com/watch?reload=9&v=MW7un0_s_Oc (zugegriffen 24.Okt.2018).
- Gutierrez, J., 2017. Son of Jaguar. YouTube. https://www.youtube.com/watch?v=e5_EE7Hpr6l

- (zugegriffen 13.Apr.2020).
- Hitchcock, A., 1948. Rope. YouTube. <https://www.youtube.com/watch?v=8xkQoH8QbVs> (zugegriffen 19.Jan.2020).
- iAnimal360, 2016. Durch die Augen eines Schweins. YouTube. https://www.youtube.com/watch?v=_pCO_mqmp6w&has_verified=1 (zugegriffen 28.Dez.2018).
- Lego, 2017. Celebrating Friendship. YouTube. https://www.youtube.com/watch?v=tAa_w4ra-fl (zugegriffen 13.Apr.2020).
- Louis Le Prince, 1888. Roundhay Garden Scene. YouTube. <https://www.youtube.com/watch?v=F1i40rnpOsA> (zugegriffen 27.Dez.2018).
- Lumière Brothers, 1895. Arrival of a Train at La Ciotat. YouTube. <https://www.youtube.com/watch?v=1dgLEDdFddk> (zugegriffen 27.Dez.2018).
- Lumière, L., 1895. La Sortie De l'Usine Lumière À Lyon. YouTube. <https://www.youtube.com/watch?v=uPmG8ppUhSw> (zugegriffen 27.Dez.2018).
- Marshmallow Laser Feast, 2015. In the Eyes of the Animal. YouTube. <https://www.youtube.com/watch?v=gfyuu01LwCE> (zugegriffen 28.Dez.2018).
- Méliès, G., 1908. Long Distance Wireless Photography. YouTube. https://www.youtube.com/watch?v=GQb_Z2RPAmS
- Méliès, G., 1902. Le Voyage dans la lune. YouTube. <https://www.youtube.com/watch?v=xLVChRVfZ74> (zugegriffen 27.Dez.2018).
- National Geographic, 2018. Galápagos 360°. YouTube. <https://www.youtube.com/watch?v=CcN6JkzM6Fk> (zugegriffen 13.Apr.2020).
- National Geographic, 2017a. Lions 360°. YouTube. <https://www.youtube.com/watch?v=sPyAQQklc1s> (zugegriffen 13.Apr.2020).
- National Geographic, 2017b. 360° Underwater National Park. YouTube. <https://www.youtube.com/watch?v=v64KOxKVLvg> (zugegriffen 13.Apr.2020).
- New York Times, 2015. The Displaced. YouTube. <https://www.youtube.com/watch?v=ecavbpCuvkl> (zugegriffen 29.Dez.2018).
- Om Nom 360°, 2017. Unexpected Guest. YouTube. https://www.youtube.com/watch?v=ct51_0Plh7s (zugegriffen 13.Apr.2020).
- Pinta Studios, 2017. The Dream Collector, YouTube. <https://www.youtube.com/watch?v=IWyBTQMwJsE> (zugegriffen 13.Apr.2020).
- Smith, G.A., 1903. The Sick Kitten. YouTube. <https://www.youtube.com/watch?v=j3u6BbA82-M> (zugegriffen 27.Dez.2018).
- Sokurow, 2002. Russian Ark. YouTube. <https://www.youtube.com/watch?v=ZV1kphEEXn8> (zugegriffen 19.Jan.2020).
- Vtopia360, 2017. Crossing Borders. vtopia360. <https://vtopia360.com/crossing-borders> (zugegriffen 9.März.2020).
- Walt Disney Studios, 2016. The Jungle Book: Through Mowgli's Eyes Part II: Cold Lairs. YouTube. <https://www.youtube.com/watch?v=bUiP-iGN6oI> (zugegriffen 13.Apr.2020).

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt ist.

München, den 23. April 2020

Sylvia Rothe